



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 108045

TO: Medina A Ibrahim
Location: cm1/9e03/9e12
Art Unit: 1638
Friday, November 14, 2003

Case Serial Number: 09/868025

From: Edward Hart
Location: Biotech-Chem Library
CM1-6B02
Phone: 305-9203

edward.hart@uspto.gov

Search Notes

Examiner Ibrahim,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart

STIC-Biotech/ChemLib

108045

RECEIVED

NOV 12 2003

(STIC)

From: Ibrahim, Medina A.
Sent: Tuesday, November 11, 2003 11:35 AM
To: STIC-Biotech/ChemLib
Subject: 09/868, 025

Please search the following:

1. DNA encoding SEQ ID NO:2.
2. SEQ ID NO:1-2.
3. Oligo of SEQ ID NO:1 and 2.

Please search both commercial and patent databases (including pending). Thanks

Medina A. Ibrahim
Patent Examiner
GAU:1638
CM1-9E03
mailbox-9E12
(703)306-5822

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 11/13/03
Date Completed: 11/17/03
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: 2
AA Sequences: 2 + 1 reverse to NA
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: 06A/04
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 13:21:00 ; Search time 30 Seconds

(without alignments)
870.200 Million cell updates/sec

Title: US-09-868-025-2

Perfect score: 803

Sequence: 1 MVDNPFISQSETHAWMS.....LGVLPKPKPEAIAKGV 143

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 102559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/ECT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/ECTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	78	9.7	257	12	US-10-179-528-4
2	75	9.3	1177	12	US-10-063-885-129
3	75	9.3	1177	15	US-10-184-644-381
4	75	9.3	1177	15	US-10-184-634-381
5	73.5	9.2	151	12	US-10-039-386-31420
6	72.5	9.0	578	14	US-10-074-527-8
7	71.5	8.9	3906	12	US-10-140-472-19
8	71.5	8.9	3906	12	US-10-141-761-19
9	71.5	8.9	3906	12	US-10-142-885-19
10	71.5	8.9	3906	12	US-10-158-790-19
11	71.5	8.9	3906	12	US-10-137-871-19
12	71.5	8.9	3906	12	US-10-140-805-19
13	71.5	8.9	3906	12	US-10-140-864-19
14	71.5	8.9	3906	12	US-10-140-923-19
15	71.5	8.9	3906	12	US-10-141-756-19

Sequence 19, Appl	71.5	8.9	3906	12	US-10-141-759-19
Sequence 19, Appl	71.5	8.9	3906	12	US-10-123-155-19
Sequence 19, Appl	71.5	8.9	3906	16	US-10-146-731-19
Sequence 17, Appl	71	8.8	131	12	US-10-231-417-174
Sequence 3, Appl	70.5	8.8	188	10	US-09-811-367B-3
Sequence 249, App	70	8.7	526	12	US-10-140-472-249
Sequence 249, App	70	8.7	526	12	US-10-141-761-249
Sequence 249, App	70	8.7	526	12	US-10-142-885-249
Sequence 249, App	70	8.7	526	12	US-10-158-790-249
Sequence 249, App	70	8.7	526	12	US-10-137-871-249
Sequence 249, App	70	8.7	526	12	US-10-140-805-249
Sequence 249, App	70	8.7	526	12	US-10-140-864-249
Sequence 249, App	70	8.7	526	12	US-10-141-756-249
Sequence 249, App	70	8.7	526	12	US-10-141-759-249
Sequence 249, App	70	8.7	526	15	US-10-123-155-249
Sequence 249, App	70	8.7	526	16	US-10-146-731-249
Sequence 3, Appl	70	8.7	1037	15	US-10-152-724A-3
Sequence 229, App	69.5	8.7	2837	12	US-10-140-472-229
Sequence 229, App	69.5	8.7	2837	12	US-10-141-761-229
Sequence 229, App	69.5	8.7	2837	12	US-10-142-885-229
Sequence 229, App	69.5	8.7	2837	12	US-10-158-790-229
Sequence 229, App	69.5	8.7	2837	12	US-10-137-871-229
Sequence 229, App	69.5	8.7	2837	12	US-10-140-805-229
Sequence 229, App	69.5	8.7	2837	12	US-10-140-864-229
Sequence 229, App	69.5	8.7	2837	12	US-10-140-923-229
Sequence 229, App	69.5	8.7	2837	12	US-10-141-756-229
Sequence 229, App	69.5	8.7	2837	12	US-10-141-759-229
Sequence 229, App	69.5	8.7	2837	15	US-10-123-155-229
Sequence 229, App	69.5	8.7	2837	16	US-10-146-731-229

ALIGNMENTS

RESULT 1

US-10-179-528-4
; Sequence 4, Application US/10179528
; Publication No. US20030166136A1
; GENERAL INFORMATION:

APPLICANT: Bandman, Olga

Shah, Purvi

TITLE OF INVENTION: NOVEL HUMAN C-TYPE LECTIN

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/179,528

FILING DATE: 24-Jun-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/846,523

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0281 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

```
/
/ LENGTH: 257 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ IMMEDIATE SOURCE:
/ LIBRARY: GenBank
/ CLONE: 505325
/ SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-10-179-528-4

Query Match          9.3%; Score 78; DB 12; Length 257;
Best Local Similarity 26.9%; Pred. No. 3.6;
Matches 35; Conservative 14; Mismatches 47; Indels 34; Gaps 7;

QY 11 QSETHAWCSSTTRSRHHHLHRRIPCLALGVTAICSLVWVHSHGGGISVALCSQ-C 69
DB 75 QSQRGSGC-----SELQRNRRRLCVALSAPV-CMLV-----LALVAVIV 113
QY 70 LQTNALRPRPD-----CLTNNGGCGEGCHGLGHVDPRPQHSNEWNSGMDCKPLRGEPLG 125
DB 114 LQRPSCSPRPFPFHVCHVNAVVGFOKCY-----YSDTESDWNSSRECHRL-GASLA 165
QY 126 VLTPHPKMEF 135
DB 166 TLDTKEMEY 175

RESULT 2
US-10-063-685-129
/ Sequence 129, Application US/10063685
/ Publication No. US20030180909A1
/ GENERAL INFORMATION:
/ APPLICANT: Eaton, Dan L.
/ APPLICANT: Filvaroff, Ellen
/ APPLICANT: Getritsen, Mary E.
/ APPLICANT: Goddard, Audrey
/ APPLICANT: Godowski, Paul J.
/ APPLICANT: Grimaldi, Christopher J.
/ APPLICANT: Gurney, Austin L.
/ APPLICANT: Watanabe, Colin K.
/ APPLICANT: Wood, William I.
/ TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
/ FILE REFERENCE: P3230R1C
/ CURRENT APPLICATION NUMBER: US/10/063.685
/ Prior Application removed - See Palm or File Wrapper
/ NUMBER OF SEQ ID NOS: 170
/ SEQ ID NO 129
/ LENGTH: 1177
/ TYPE: DNA
/ ORGANISM: Homo Sapien
US-10-063-685-129

Query Match          9.3%; Score 75; DB 12; Length 1177;
Best Local Similarity 32.1%; Pred. No. 43;
Matches 18; Conservative 1; Mismatches 21; Indels 16; Gaps 2;

QY 39 CLALGVTAICSLVWVHSHGGGISVALCSQCLQTNALRPRPDCLTNNGGCGEGCHG 94
DB 114 CTCAGTTGCC-----GGAGGACCTCCCT-----CCCGAGGCGAGTCTG 153

RESULT 3
US-10-184-644-381
/ Sequence 381, Application US/10184644
/ Publication No. US20030044930A1
/ GENERAL INFORMATION:
/ APPLICANT: Baker, Kevin P.
/ APPLICANT: Chen, Jian
/ APPLICANT: Desnoyers, Luc
/ APPLICANT: Goddard, Audrey
/ APPLICANT: Godowski, Paul J.
/ TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
/ FILE REFERENCE: P3430R1C217
/ CURRENT APPLICATION NUMBER: US/10/184.634
/ Prior Application removed - See File Wrapper or Palm
/ NUMBER OF SEQ ID NOS: 612
/ SEQ ID NO 381
/ LENGTH: 1177
/ TYPE: DNA
/ ORGANISM: Homo Sapien
US-10-184-634-381

Query Match          9.3%; Score 75; DB 15; Length 1177;
Best Local Similarity 32.1%; Pred. No. 43;
Matches 18; Conservative 1; Mismatches 21; Indels 16; Gaps 2;

QY 39 CLALGVTAICSLVWVHSHGGGISVALCSQCLQTNALRPRPDCLTNNGGCGEGCHG 94
DB 114 CTCAGTTGCC-----GGAGGACCTCCCT-----CCCGAGGCGAGTCTG 153

RESULT 5
US-10-029-386-31420
/ Sequence 31420, Application US/10029386
/ Publication No. US20030194704A1
/ GENERAL INFORMATION:
/ APPLICANT: Goddard, Audrey
/ APPLICANT: Penn, Sharon G.
/ APPLICANT: Rank, David R.
```

```
/
/ APPLICANT: Gurney, Austin L.
/ APPLICANT: Pan, James
/ APPLICANT: Smith, Victoria
/ APPLICANT: Watanabe, Colin K.
/ APPLICANT: Wood, William I.
/ APPLICANT: Zhang, Zemin
/ TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
/ FILE REFERENCE: P3430R1C227
/ CURRENT APPLICATION NUMBER: US/10/184.644
/ Prior Application removed - See File Wrapper or Palm
/ NUMBER OF SEQ ID NOS: 612
/ SEQ ID NO 381
/ LENGTH: 1177
/ TYPE: DNA
/ ORGANISM: Homo Sapien
US-10-184-644-381

Query Match          9.3%; Score 75; DB 15; Length 1177;
Best Local Similarity 32.1%; Pred. No. 43;
Matches 18; Conservative 1; Mismatches 21; Indels 16; Gaps 2;

QY 39 CLALGVTAICSLVWVHSHGGGISVALCSQCLQTNALRPRPDCLTNNGGCGEGCHG 94
DB 114 CTCAGTTGCC-----GGAGGACCTCCCT-----CCCGAGGCGAGTCTG 153

RESULT 4
US-10-184-634-381
/ Sequence 381, Application US/10184634
/ Publication No. US20030068684A1
/ GENERAL INFORMATION:
/ APPLICANT: Baker, Kevin P.
/ APPLICANT: Chen, Jian
/ APPLICANT: Desnoyers, Luc
/ APPLICANT: Goddard, Audrey
/ APPLICANT: Godowski, Paul J.
/ APPLICANT: Gurney, Austin L.
/ APPLICANT: Pan, James
/ APPLICANT: Smith, Victoria
/ APPLICANT: Watanabe, Colin K.
/ APPLICANT: Wood, William I.
/ APPLICANT: Zhang, Zemin
/ TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
/ FILE REFERENCE: P3430R1C217
/ CURRENT APPLICATION NUMBER: US/10/184.634
/ Prior Application removed - See File Wrapper or Palm
/ NUMBER OF SEQ ID NOS: 612
/ SEQ ID NO 381
/ LENGTH: 1177
/ TYPE: DNA
/ ORGANISM: Homo Sapien
US-10-184-634-381

Query Match          9.3%; Score 75; DB 15; Length 1177;
Best Local Similarity 32.1%; Pred. No. 43;
Matches 18; Conservative 1; Mismatches 21; Indels 16; Gaps 2;

QY 39 CLALGVTAICSLVWVHSHGGGISVALCSQCLQTNALRPRPDCLTNNGGCGEGCHG 94
DB 114 CTCAGTTGCC-----GGAGGACCTCCCT-----CCCGAGGCGAGTCTG 153

RESULT 5
US-10-029-386-31420
/ Sequence 31420, Application US/10029386
/ Publication No. US20030194704A1
/ GENERAL INFORMATION:
/ APPLICANT: Goddard, Audrey
/ APPLICANT: Penn, Sharon G.
/ APPLICANT: Rank, David R.
```


APPLICANT: Harzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: AEMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annotation Sequence Listing Engine vers. 1.1
; SEQ ID NO 31420
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005329.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.81
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
US-10-029-386-31420

Query Match 9.2%; Score 73.5; DB 12; Length 151;
Best Local Similarity 25.7%; Pred. No. 5.9;
Matches 35; Conservative 12; Mismatches 42; Indels 47; Gaps 7;

QY 19 WSS-----STTRSPSRHHRERIPCLALGVTAICSLVWIVHSHGGGISVALCS 67
DB 48 WTSVFTDLCTCLSTPRAGNPAVAMQRLPCPHAGV-----CLVWSSGALPGGGGGGGC- 102
QY 68 QCLQTNALRPRPDLCTNNGCGYCHGSLGHVDRPQHSNWSGMDSC--CKPLRGHFLG 125
DB 103 -----RWPPPA-----GRCFGDSGRVPVPPPE-----AGLSSVAC-----TG 133
QY 126 VLTPHKWEFAAIRAG 141
DB 134 LVLPHGKSSGAALRG 149

RESULT 6
US-10-074-527-8
; Sequence 8, Application US/10074527
; Publication No. US20020142426A1
; GENERAL INFORMATION:
; APPLICANT: Olandt, Peter J.
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Galvin, Katherine A.
; APPLICANT: Millennium Pharmaceuticals Inc.
; TITLE OF INVENTION: 33945, A Human Glycosyltransferase and
; TITLE OF INVENTION: Uses Therefor
; FILE REFERENCE: MPI2001-018PACF1(M)
; CURRENT APPLICATION NUMBER: US/10/074,527
; CURRENT FILING DATE: 2002-02-12
; PRIOR APPLICATION NUMBER: 60/269202
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 578
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-074-527-8

Query Match 9.0%; Score 72.5; DB 14; Length 578;
Best Local Similarity 23.4%; Pred. No. 35;
Matches 37; Conservative 20; Mismatches 46; Indels 55; Gaps 9;

QY 28 SRHHLRERIPCLALGVTAICSLVW-----IHVS-----HGGGISVALCSQCCLQTN 73
DB 410 SERKLLRERLACK-----SFDWYLNQVPSNLHLVPEDRPGWGAIRSMGISSECLDYN 461
QY 74 ALRPRPDLCTNNGCGYCHGSLGHVDRPQHSNE---WNSGMDSC--PLRGHFLG--- 125
DB 462 A--PONNPTGANLSLFG--CHQGGN--QPFVTSNKEIRFNSTELCAEVPOQKQIVGMQN 517
QY 126 -----VLTPHKWEFAAIRAG 142

DB 518 CPKDLGVPEVNIWIKEDGTIFPHPTRLCLLSAYRTAE 555

RESULT 7
US-10-140-472-19
; Sequence 19, Application US/10140472
; Publication No. US2003013888A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C198
; CURRENT APPLICATION NUMBER: US/10/140,472
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 19
; LENGTH: 3906
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-140-472-19

Query Match 8.9%; Score 71.5; DB 12; Length 3906;
Best Local Similarity 30.5%; Pred. No. 3.9e+02;
Matches 18; Conservative 2; Mismatches 32; Indels 7; Gaps 1;

QY 39 CLALGVTAICSLVWIVHSHGGGISVALCSQCCLQTNALRPRPDLCTNNGCGYCHGSLG 97
DB 713 CAGAGGGACGCG-----AGGACATCGCCCCACATCATCTCTGCTACGGGACG 764

RESULT 8
US-10-141-761-19
; Sequence 19, Application US/10141761
; Publication No. US20030148432A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C198
; CURRENT APPLICATION NUMBER: US/10/141,761

; CURRENT FILING DATE: 2002-05-08
 ; Prior Application removed - See Palm or File Wrapper
 ; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 19
 ; LENGTH: 3906
 ; TYPE: DNA
 ; ORGANISM: Homo Sapien
 US-10-141-761-19

Query Match 8.9%; Score 71.5; DB 12; Length 3906;
 Best Local Similarity 30.5%; Pred. No. 3.9e+02;
 Matches 18; Conservative 2; Mismatches 32; Indels 7; Gaps 1;

QY 39 CLALGVTAICSLVWVHSHGGISVALCSQCLQTNALRPRDCLTNNGCYGEGCHGSLG 97
 DB 713 CAGAGGGACGCG-----AGGAACATCCGCCCAACATCATCTCGTGTGCGAGCG 764

RESULT 9

US-10-142-885-19
 ; Sequence 19, Application US/10142885
 ; Publication No. US20030157604A1

GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.
 ; APPLICANT: Beresini, Maureen
 ; APPLICANT: Deforge, Laura
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Filvaroff, Ellen
 ; APPLICANT: Gao, Wei-Qiang
 ; APPLICANT: Gerritsen, Mary E.
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Sherwood, Steven
 ; APPLICANT: Smith, Victoria
 ; APPLICANT: Stewart, Timothy A.
 ; APPLICANT: Tumas, Daniel
 ; APPLICANT: Watanabe, Colin K
 ; APPLICANT: Wood, William
 ; APPLICANT: Zhang, Zemin
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; FILE REFERENCE: P3330R1C248
 ; CURRENT APPLICATION NUMBER: US/10/142.885
 ; CURRENT FILING DATE: 2002-05-10
 ; Prior Application removed - See File Wrapper or Palm
 ; NUMBER OF SEQ ID NOS: 550
 ; SEQ ID NO 19
 ; LENGTH: 3906
 ; TYPE: DNA
 ; ORGANISM: Homo Sapien
 US-10-142-885-19

Query Match 8.9%; Score 71.5; DB 12; Length 3906;
 Best Local Similarity 30.5%; Pred. No. 3.9e+02;
 Matches 18; Conservative 2; Mismatches 32; Indels 7; Gaps 1;

QY 39 CLALGVTAICSLVWVHSHGGISVALCSQCLQTNALRPRDCLTNNGCYGEGCHGSLG 97
 DB 713 CAGAGGGACGCG-----AGGAACATCCGCCCAACATCATCTCGTGTGCGAGCG 764

RESULT 10

US-10-158-790-19
 ; Sequence 19, Application US/10158790
 ; Publication No. US20030180879A1

GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.
 ; APPLICANT: Beresini, Maureen
 ; APPLICANT: Deforge, Laura
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Filvaroff, Ellen
 ; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Sherwood, Steven
 ; APPLICANT: Smith, Victoria
 ; APPLICANT: Stewart, Timothy A.
 ; APPLICANT: Tumas, Daniel
 ; APPLICANT: Watanabe, Colin K
 ; APPLICANT: Wood, William
 ; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; FILE REFERENCE: P3330R1C448
 ; CURRENT APPLICATION NUMBER: US/10/158.790
 ; CURRENT FILING DATE: 2002-05-30
 ; Prior Application removed - See File Wrapper or Palm
 ; NUMBER OF SEQ ID NOS: 550
 ; SEQ ID NO 19
 ; LENGTH: 3906
 ; TYPE: DNA
 ; ORGANISM: Homo Sapien
 US-10-158-790-19

Query Match 8.9%; Score 71.5; DB 12; Length 3906;

Best Local Similarity 30.5%; Pred. No. 3.9e+02;
Matches 18; Conservative 2; Mismatches 32; Indels 7; Gaps 1;

QY 39 CLALGVTAICSLVWVHSHGGISVALCSQCLQTNALRPRDCLTNNGCYGEGCHGSLG 97
 DB 713 CAGAGGGACGCG-----AGGAACATCCGCCCAACATCATCTCGTGTGCGAGCG 764

RESULT 11

US-10-137-871-19
 ; Sequence 19, Application US/10137871
 ; Publication No. US20030207350A1

GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.
 ; APPLICANT: Beresini, Maureen
 ; APPLICANT: Deforge, Laura
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Filvaroff, Ellen
 ; APPLICANT: Gao, Wei-Qiang
 ; APPLICANT: Gerritsen, Mary E.
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Sherwood, Steven
 ; APPLICANT: Smith, Victoria
 ; APPLICANT: Stewart, Timothy A.
 ; APPLICANT: Tumas, Daniel
 ; APPLICANT: Watanabe, Colin K
 ; APPLICANT: Wood, William
 ; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; FILE REFERENCE: P3330R1C153

; CURRENT APPLICATION NUMBER: US/10/137.871
 ; CURRENT FILING DATE: 2002-05-03

; Prior Application removed - See Palm or File Wrapper
 ; NUMBER OF SEQ ID NOS: 550
 ; SEQ ID NO 19
 ; LENGTH: 3906
 ; TYPE: DNA
 ; ORGANISM: Homo Sapien
 US-10-137-871-19

Query Match 8.9%; Score 71.5; DB 12; Length 3906;

Best Local Similarity 30.5%; Pred. No. 3.9e+02;
Matches 18; Conservative 2; Mismatches 32; Indels 7; Gaps 1;

QY 39 CLALGVTAICSLVWVHSHGGISVALCSQCLQTNALRPRDCLTNNGCYGEGCHGSLG 97


```

; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tamas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330RIC200
; CURRENT APPLICATION NUMBER: US/10/141,756
; CURRENT FILING DATE: 2002-05-08
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 19
; LENGTH: 3906
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-141-756-19

Query Match      8.9%; Score 71.5; DB 12; Length 3906;
Best Local Similarity 30.5%; Pred. No. 3.9e+02;
Matches 18; Conservative 2; Mismatches 32; Indels 7; Gaps 1;

QY      39 CLALGVTAICSLVWVHSHGGGIVSVALCSQCLQTNALRPDPCLTNNGGCGYGECHGSLG 97
DB      713 CAGAGGACCGC-----AGGACATCCGCCCAACATCATCTCTGTGTGTGACGACG 764

```

Search completed: November 13, 2003, 13:26:27
Job time : 30 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 13:18:44 ; Search time 35 Seconds
(without alignments)
1054.329 Million cell updates/sec

Title: US-09-868-025-2
Perfect score: 803
Sequence: 1 MVDNFPISQSETHAWCHS.....LGVLPHPKMEFAAIRAGKV 143

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vestibate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	803	100.0	143	10 Q9SNZ5	Q9snz5 oryza sativ
2	82.5	10.3	431	10 Q9L1N4	Q9l1n4 oryza sativ
3	78	9.7	257	13 Q90636	Q90636 gallus gall
4	77.5	9.7	930	16 Q9KFM7	Q9kfm7 bacillus ha
5	76	9.5	110	10 Q9AVS3	Q9avb3 pisum sativ
6	74.5	9.3	154	11 Q8BHM0	Q8bhm0 mus musculu
7	74.5	9.3	210	11 Q8BT65	Q8bt65 mus musculu
8	74.5	9.3	995	11 Q8BR76	Q8br76 mus musculu
9	74	9.2	240	4 Q9H4A7	Q9h4a7 homo sapien
10	73.5	9.2	420	4 O15470	O15470 homo sapien
11	73.5	9.2	579	5 Q9N4P6	Q9n4p6 caenorhabdi
12	73.5	9.2	1637	6 Q9XSV8	Q9xsv8 bos taurus
13	73.5	9.2	5146	6 Q8SPM4	Q8spm4 bos taurus
14	73	9.1	310	4 Q8WYX3	Q8wyx3 homo sapien
15	72.5	9.0	578	11 O08832	O08832 mus musculu
16	72.5	9.0	4998	11 Q8CG65	Q8cgg65 mus musculu

17	72	9.0	701	11 Q8K4K9	Q8k4k9 rattus norv
18	72	9.0	764	10 Q9L1M7	Q9l1m7 arabidopsis
19	72	9.0	872	5 Q22580	Q22580 caenorhabdi
20	72	9.0	1664	5 Q9TVQ2	Q9tvq2 caenorhabdi
21	71.5	8.9	208	10 Q8S0Z1	Q8s0z1 oryza sativ
22	71.5	8.9	429	16 Q8FD12	Q8fd12 escherichia
23	71.5	8.9	1001	10 Q8LSQ6	Q8lsq6 oryza sativ
24	71	8.8	138	12 Q68244	Q68244 hepatitis c
25	71	8.8	138	12 Q68205	Q68205 hepatitis c
26	71	8.8	322	16 Q8FVR0	Q8fvr0 brucella su
27	71	8.8	325	16 Q8YCK2	Q8yck2 brucella me
28	71	8.8	337	10 Q9SH41	Q9sh41 arabidopsis
29	71	8.8	3010	12 Q9Q1Y1	Q9q1y1 hepatitis c
30	71	8.8	3010	12 Q93016	Q93016 hepatitis c
31	71	8.8	3010	12 Q9Q1Y2	Q9q1y2 hepatitis c
32	71	8.8	3012	12 Q9WIK7	Q9wik7 hepatitis c
33	70.5	8.8	188	11 Q88713	Q88713 mus musculu
34	70.5	8.8	429	16 Q8XBE5	Q8xeb5 escherichia
35	70.5	8.8	537	4 Q9HBE2	Q9hbe2 homo sapien
36	70.5	8.8	641	11 Q9JMG9	Q9jmg9 mus musculu
37	70.5	8.8	1376	5 Q8S2S2	Q8s2s2 drosophila
38	70	8.7	133	17 Q8ZM63	Q8zm63 pyrobaculum
39	70	8.7	543	10 Q8LHK6	Q8lhx6 oryza sativ
40	70	8.7	1028	11 Q9JLL0	Q9jll0 mus musculu
41	70	8.7	3010	12 Q9Q1X1	Q9q1x1 hepatitis c
42	70	8.7	3010	12 Q9Q1X3	Q9q1x3 hepatitis c
43	70	8.7	3010	12 Q9Q1X4	Q9q1x4 hepatitis c
44	70	8.7	3010	12 Q9DTE5	Q9dte5 hepatitis c
45	69	8.6	138	12 Q68231	Q68231 hepatitis c

ALIGNMENTS

RESULT 1

Q9SNZ5 PRELIMINARY, PRT, 143 AA.
AC Q9SNZ5;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Hypothetical 15.7 kDa protein (Fragment).
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. IR64; TISSUE=Seedling;
RA Patel V.M., Machai C.A., Divya C., Ashok M.;
RT "Oryza sativa Variety IR64 (cDNA clone AGRSAL-11 from 7 days old seedling).";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF192975; AAF06789.1; -
KW Gramine; Q9SNZ5; -
DR Hypothetical protein.
FT NON_TER 143 143
SQ SEQUENCE 143 AA; 15689 MW; 1CD55C1EFA9B0AD8 CRC64;

Query Match 100.0%; Score 803; DB 10; Length 143;
Best Local Similarity 100.0%; Pred. No. 7.9e-83;
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDNFPISQSETHAWCHSSTTRSPSRHLHRRIPCLALGYTAICSLVHVSHGCG 60
DB 1 MVDNFPISQSETHAWCHSSTTRSPSRHLHRRIPCLALGYTAICSLVHVSHGCG 60
QY 61 ISVALCSQCLTNALRPPDCLTNGGCGYCHGSLGHVDFPQHSNENWNSGMDCKPLR 120
DB 61 ISVALCSQCLTNALRPPDCLTNGGCGYCHGSLGHVDFPQHSNENWNSGMDCKPLR 120
QY 121 GEFLOVLTTPHPKMEFAAIRAGKV 143

```

Db      121 GEFLOVLTTPHPKMEFAAIRACKV 143
|||||
Q9LLN4 PRELIMINARY; PRT; 431 AA.
AC Q9LLN4
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical 49.0 kDa protein.
GN DUPR11.22.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Euphorbiaceae; Oryzae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=20181732; PubMed=10715324;
RX Tarchini R., Biddle P., Wineland R., Tingey S., Rafalski A.;
RT "The complete sequence of 340 kb of DNA around the rice Adh1-adh2
RT region reveals interrupted colinearity with maize chromosome 4.";
RL Plant Cell 12:381-391(2000).
DR EMBL; AF172282; AAF34429.1; -.
DR Gramineae; O9LLN4; -.
DR InterPro; IPR001810; F-box.
DR Pfam; PF00646; F-box; 1.
DR SMART; SM00256; FBOX; 1.
DR PROSITE; PS50181; FBOX; 1.
KW Hypothetical protein.
SQ SEQUENCE 431 AA; 49012 MW; FD22807C60D963B9 CRC64;

Query Match 10.3%; Score 82.5; DB 10; Length 431;
Best Local Similarity 26.7%; Pred. No. 0.65;
Matches 39; Conservative 10; Mismatches 38; Indels 59; Gaps 8;

QY 5 NFP-ISEQETHAW-----CWSSSTTRSPSRHHLRERIPCLALGVTAICSLVW 52
DB 194 DFEFSEHURTHIYAPNLITLDDCG-----RVFFLESMPSLIAFVR 238
QY 53 IHVSHGGGTSVALCSQCLOTNALRPRDCLTNNGGCYGE-CHGSLGHVDRFPQHSNWN 111
DB 239 PHRD-----SDLCNSN-----TSGNCEDEYCHGCGMVD-----NA 270
QY 112 GNDCKPLRGEFLVLTTPHPKMEFA 137
DB 271 GNDACKV---LLGLSEAKKLELIA 293

RESULT 3
Q90636 PRELIMINARY; PRT; 257 AA.
AC Q90636
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE 17.5 protein.
GN 17.5.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Spleen;
RX MEDLINE=94164691; PubMed=8119728;
RA Bernot A., Zoorob R., Auffray C.;
RT "Linkage of a new member of the lectin supergene family to the chicken
RT Mhc genes.";
RL Immunogenetics 39:221-229(1994).

```

```

DR EMBL; M88072; AAA48558.1; -.
DR InterPro; IPR002353; AntifreezeZell.
DR InterPro; IPR001304; Lectin_C.
DR Pfam; PF00059; Lectin_C; 1.
DR PRINTS; PR00356; ANTIFREEZEII.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS0041; C_TYPE_LLECTIN_2; 1.
SQ SEQUENCE 257 AA; 28814 MW; C67EA483F4E85333 CRC64;

Query Match 9.7%; Score 78; DB 13; Length 257;
Best Local Similarity 26.9%; Pred. No. 1.2;
Matches 35; Conservative 14; Mismatches 47; Indels 34; Gaps 7;

QY 11 QSETHAWSSSTTRSPSRHHLRERIPCLALGVTAICSLVWIHVSHGGGTSVALCSQ-C 69
DB 75 OSQSGSGC-----SELQRNRREVLCAVSAVP-CMLV-----LALVAVIV 113
QY 70 LQTNALRPRD-----CLTNNGGCYGECHGSLGHVDRFPQHSNWNWMDCKPLRGEFLG 125
DB 114 LQRPSCSPRPSPSHVCPNNAWVGQKCY-----YFSDTESDWNSSREHCHRL-GASLA 165
QY 126 VLTPHPKMEF 135
DB 166 TLDTKSEMEF 175

RESULT 4
Q9KFM7 PRELIMINARY; PRT; 990 AA.
AC Q9KFM7
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Lanthibiotic mercacidin modifying enzyme.
GN BH0452.
OC Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirata C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AP001508; BA804171.1; -.
KW Complete proteome.
SQ SEQUENCE 990 AA; 112397 MW; 73B556D2A0D92236 CRC64;

Query Match 9.7%; Score 77.5; DB 16; Length 990;
Best Local Similarity 23.6%; Pred. No. 5.8;
Matches 25; Conservative 12; Mismatches 38; Indels 31; Gaps 3;

QY 52 WIHVSHGGGTSVALCSQCLOTNALRPRDCLTN-----GGCYGECHGSLGHV-----RF 102
DB 847 WCHGAPGIGISRLLLAOFYDELLHEELNAALNTISDGFHNSLCHGDFGNLDDLLLY 906
QY 103 POHSGNE-----WNSGMDCKPLRGEFLGV 126
DB 907 AQYTNPEPKELARKLAISSIDQATYGVKGLNHSQLOGMMLGV 952

RESULT 5
Q9AVS3 PRELIMINARY; PRT; 110 AA.
AC Q9AVS3
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Trypsin inhibitor precursor.

```

Db 122 GHI 124

RESULT 7

Q8BT65 PRELIMINARY; PRT; 210 AA.

ID Q8BT65 AC Q8BT65; TISSUE=Brain;

DT 01-MAR-2003 (T-EMBLrel. 23, Created)

DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)

DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)

DE Hypothetical protein.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OC NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RP STRAIN=C57BL/6J; TISSUE=Brain;

RC MEDLINE=22354683; PubMed=12466851;

RA The FANTOM Consortium,

RA the RIKEN Genome Exploration Research Group Phase I & II Team;

RT "Analysis of the mouse transcriptome based on functional annotation of

RT 60,770 full-length cDNAs.";

RL Nature 420:563-573(2002).

DR EMBL: AK017238; BAC25507.1; -.

KW Hypothetical protein.

SQ SEQUENCE 210 AA; 22692 MW; 8D7197233A1E7EF8 CRC64;

Query Match 9.3%; Score 74.5; DB 11; Length 210;

Best Local Similarity 28.6%; Pred. No. 2.4;

Matches 18; Conservative 14; Mismatches 26; Indels 5; Gaps 3;

QY 41 ALGVTAICSLVTHVSHGGGISVALCSQCLOT--NALRPRDCLTNNGG--CYGECHGSL 96

DB 63 ALGTSVCVCLPGYHMINNGGPSI-ICKKCPENMKGVTKDGMDCISCPSGLTAEGKCHCPT 121

QY 97 GHV 99

DB 122 GHI 124

RESULT 8

Q8BT76 PRELIMINARY; PRT; 995 AA.

ID Q8BT76 AC Q8BT76; TISSUE=Brain;

DT 01-MAR-2003 (T-EMBLrel. 23, Created)

DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)

DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)

DE Hypothetical protein.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OC NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RP STRAIN=C57BL/6J; TISSUE=Brain;

RC MEDLINE=22354683; PubMed=12466851;

RA The FANTOM Consortium,

RA the RIKEN Genome Exploration Research Group Phase I & II Team;

RT "Analysis of the mouse transcriptome based on functional annotation of

RT 60,770 full-length cDNAs.";

RL Nature 420:563-573(2002).

DR EMBL: AK045429; BAC32362.1; -.

KW Hypothetical protein.

SQ SEQUENCE 995 AA; 112099 MW; ECC13755D91197CB CRC64;

Query Match 9.3%; Score 74.5; DB 11; Length 995;

Best Local Similarity 28.6%; Pred. No. 13;

Matches 18; Conservative 14; Mismatches 26; Indels 5; Gaps 3;

QY 41 ALGVTAICSLVTHVSHGGGISVALCSQCLOT--NALRPRDCLTNNGG--CYGECHGSL 96

DB 76 ALGTSVCVCLPGYHMINNGGPSI-ICKKCPENMKGVTKDGMDCISCPSGLTAEGKCHCPT 134

```

QY 97 GHV 99
DB 135 GHI 137

RESULT 9
Q9H4A7 PRELIMINARY; PRT; 260 AA.
AC Q9H4A7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Synapcobrevin like protein 1B.
GN SYBLB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1];
RP SEQUENCE FROM N.A.
RA D'Esposito M., Filippini F., Rossi V., D'Urso M.;
RT "Alternative splicing of SYBL1 gene.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ295938; CAC16891.1;
DR InterPro; IPR001388; Synapcobrevin.
DR Pfam; PF00957; synapcobrevin; 1.
SQ SEQUENCE 260 AA; 30217 MW; AB6942EF4ACD2978 CRC64;

Query Match 9.2%; Score 74; DB 4; Length 260;
Best Local Similarity 27.7%; Pred. No. 3.5;
Matches 23; Conservative 9; Mismatches 41; Indels 10; Gaps 3;

QY 20 SSTRSPSRHLHRRIPCLALGVTAICSLVWIVHSHGGGISVALCSQCILQTNALPRP 79
DB 157 SSHVVEEQAYHHRRINCVEL-YHCFTSLWIIYA-----KICEICKKL--PLT 206

QY 80 DCLTNGGCGEGCHOSLGHVDRF 102
DB 207 KDMREQGVKSNPCDSLSLHTDRA 229

RESULT 10
Q15470 PRELIMINARY; PRT; 420 AA.
AC Q15470;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Monocyte inhibitory receptor precursor (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1];
RP SEQUENCE FROM N.A.
RA Arm J.P., Nwankwo C., Austen K.F.;
RT "Molecular identification of a novel family of human immunoglobulin
superfamily members that possess immunoreceptor tyrosine-based
inhibitory motifs and homology to the mouse gp49BI inhibitory
receptor.";
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U91927; AAB68667.1;
DR HSSP; P43626; INKR.
DR InterPro; IPR003599; IG.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; IG; 4.
DR SMART; SM00409; IG; 3.
DR PROSITE; PS00835; IG_LIKE; 3.
KW Receptor; Signal.
FT NON_TER 1 1 4
FT SIGNAL <1 4 POTENTIAL.

FT CHAIN 5 420 MONOCYTE INHIBITORY RECEPTOR.
SQ SEQUENCE 420 AA; 45469 MW; 416301823B6433B8 CRC64;

Query Match 9.2%; Score 73.5; DB 4; Length 420;
Best Local Similarity 22.1%; Pred. No. 6.6;
Matches 45; Conservative 20; Mismatches 64; Indels 75; Gaps 10;

QY 8 ISEQSETHAWSSSTTRSPSRHLHRE-----RIP--CLALGVTAICSLVWIIH--- 54
DB 20 ITQSPVTLRCQGSLETQ---EYHLYREKKTALWITRIFQELVKKGQFFILSITWEHAGR 76
QY 55 -----VSHGGGISVALCSQ-----C 69
DB 77 YCCIYGSHTAGLSSESDPLELVVTGAYSKPTLSALPSPVVTSGENVTIQCDQVAFDGF 136
QY 70 LQTNALRPRPDCLTNGGCGEGCHG--SLGHVDRFPQHSNEWN---SGMDSKFP-----LR 120
DB 137 LCKEGEDEHPQCLNHHARGSSRAIFSYGPV-----SPSRMSYRCYGYDSRAPYVWSLP 192
QY 121 GEFLGVLTGH-PKMEPAATRAKGV 143
DB 193 SDLLGLLVPGVPKPKSLSVQPGPV 216

RESULT 11
Q9N4P6 PRELIMINARY; PRT; 579 AA.
AC Q9N4P6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical 66.6 kDa protein.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Paloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1];
RP SEQUENCE FROM N.A.
RA STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018 (1998).
RN [2];
RP SEQUENCE FROM N.A.
RA STRAIN=Bristol N2;
RT "The sequence of C. elegans cosmid Y81B9A.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3];
RP SEQUENCE FROM N.A.
RA STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006827; AAP60881.1;
DR WormPep; Y81B9A.1; CE231116.
KW Hypothetical protein.
SQ SEQUENCE 579 AA; 66614 MW; 039D418555213C10 CRC64;

Query Match 9.2%; Score 73.5; DB 5; Length 579;
Best Local Similarity 30.4%; Pred. No. 9.3;
Matches 21; Conservative 12; Mismatches 23; Indels 13; Gaps 5;

QY 30 HHLHRRIPCLALGVTAICSLVWIVHSHGGGISVALCSQCILQTNALRPRPDCLTNGGCGY 89
DB 6 HKLH---DCVFNKNTTMELEFMAVNHV-----TYCGGCTNRSHMPRGEC--QRVGCY 54
QY 90 GECHGSLGH 98
DB 55 -QC-GSMGH 61

```



```

12
JXSV8      PRELIMINARY;      PRT; 1637 AA.
J1-NOV-1999 (TRENBLrel. 12, Created)
J1-NOV-1999 (TRENBLrel. 12, Last sequence update)
01-MAR-2003 (TRENBLrel. 23, Last annotation update)
SCO-spandin (Fragment).
SCO-SPONDIN.
Bos taurus (Bovine).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
NCBI_TaxID=9913;
[1]
SEQUENCE FROM N.A.
Gobron S.; Creveaux I.; Didier R.; Meinzel R.;
"Characterization of cattle SCO-spandin."
Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
EMBL; AJ133488; CAB46213.1; -.
HSP; P56682; ICCV.
InterPro: IPR006207; Cys_knot.
InterPro: IPR006207; Cys_knot_C.
InterPro: IPR001545; Gly_hormones.
InterPro: IPR002919; TIL_Cyrich.
InterPro: IPR000884; TSPI.
InterPro: IPR001007; VWF_C.
Pfam; PF01826; TIL; 10.
Pfam; PF00090; TSP; 1; 25.
Pfam; PF00094; VWF; 3.
SMART; SM00041; CT; 1.
SMART; SM00068; GHB; 1.
SMART; SM00192; LDLA; 10.
SMART; SM00209; TSP1; 25.
SMART; SM00216; VWD; 3.
PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
PROSITE; PS01225; CTCK; 2; 1.
PROSITE; PS01285; FAS8C; 1; 1.
PROSITE; PS01286; FAS8C; 2; 1.
PROSITE; PS00261; GLYCO_HORMONE_BETA_1; 2.
PROSITE; PS01209; LDLA; 1; 7.
PROSITE; PS00668; LDLA; 2; 9.
PROSITE; PS00092; TSP1; 25.
PROSITE; PS01208; VWF; 1.
SQ
SEQUENCE 1637 AA; 173347 MW; 4CSB81DD346C925 CRC64;
Query Match 9.24; Score 73.5; DB 6; Length 1637;
Best Local Similarity 24.64; Pred. No. 28;
Matches 34; Conservative 11; Mismatches 44; Indels 49; Gaps 9;
QY 33 HRRIP-----CLALGVTAICSL-----VMIHSHGGGISVALCSQ-CLQ 71
Db 1380 HRELPPTLLTQNTCHICCGAFSCSLTDCQCPPTWQVAPG---ELGECROTCE 1436
QY 72 TNAL-----RPRDCL-----TNGGCVG-----ECHGSLGHVDRFPQHSNENSG 112
Db 1437 PNATETQNCGRQAPGVQCGVQRGHRPSQEGPCVPDLCEC-----WHGRPHPPGSEWQKA 1492
QY 113 MDSCKPLRGEFLGVLTTPH 130
Db 1493 CESRCVSGE--SICQH 1508
RESULT 13
Q8SPM4      PRELIMINARY;      PRT; 5146 AA.
AC Q8SPM4;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE SCO-spandin.
GN SCO-SPONDIN.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC "a.a. Bovinae; Bos.
[1]
SEQUENCE FROM N.A.
Tissue-Subcommittee organ;
Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.
Tissue-Subcommittee organ;
MEDLINE=20465125; PubMed=11008217;
Gobron S.;
"Subcommittee organ/Reissner's fiber complex: characterization of
SCO-spandin, a glycoprotein with potent activity on neurite
outgrowth."
Glia 32:177-181 (2000).
EMBL; AJ16457; CAC94914.1; -.
InterPro: IPR001064; Crystallin.
InterPro: IPR006208; Cys_knot.
InterPro: IPR006207; Cys_knot_C.
InterPro: IPR000421; FAS8_C.
InterPro: IPR001545; Gly_hormones.
InterPro: IPR002172; LDL_receptor_A.
InterPro: IPR002319; TIL_Cyrich.
InterPro: IPR000884; TSPI.
InterPro: IPR001007; VWF_C.
InterPro: IPR001846; VWF_D.
Pfam; PF00007; Cys_knot; 1.
Pfam; PF00754; F5_P8_type_C; 1.
Pfam; PF00057; LDL_receptor_a; 10.
Pfam; PF01826; TIL; 10.
Pfam; PF00090; TSP; 1; 25.
Pfam; PF00094; VWF; 3.
SMART; SM00041; CT; 1.
SMART; SM00231; FAS8C; 1.
SMART; SM00068; GHB; 1.
SMART; SM00192; LDLA; 10.
SMART; SM00209; TSP1; 25.
SMART; SM00216; VWD; 3.
PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
PROSITE; PS01225; CTCK; 2; 1.
PROSITE; PS01285; FAS8C; 1; 1.
PROSITE; PS01286; FAS8C; 2; 1.
PROSITE; PS00261; GLYCO_HORMONE_BETA_1; 2.
PROSITE; PS01209; LDLA; 1; 7.
PROSITE; PS00668; LDLA; 2; 9.
PROSITE; PS00092; TSP1; 25.
PROSITE; PS01208; VWF; 1.
SQ
SEQUENCE 5146 AA; 543576 MW; 724CSFB6727E13DA CRC64;
Query Match 9.24; Score 73.5; DB 6; Length 5146;
Best Local Similarity 24.64; Pred. No. 96;
Matches 34; Conservative 11; Mismatches 44; Indels 49; Gaps 9;
QY 33 HRRIP-----CLALGVTAICSL-----VMIHSHGGGISVALCSQ-CLQ 71
Db 4889 HRELPPTLLTQNTCHICCGAFSCSLTDCQCPPTWQVAPG---ELGECROTCE 4945
QY 72 TNAL-----RPRDCL-----TNGGCVG-----ECHGSLGHVDRFPQHSNENSG 112
Db 4946 PNATETQNCGRQAPGVQCGVQRGHRPSQEGPCVPDLCEC-----WHGRPHPPGSEWQKA 5001
QY 113 MDSCKPLRGEFLGVLTTPH 130
Db 5002 CESRCVSGE--SICQH 5017
RESULT 14
Q8WYX3      PRELIMINARY;      PRT; 310 AA.
AC Q8WYX3;
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Hypothetical protein.

```

GN P21741.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCES FROM N.A.
RA Zhang P.P., Zhou X.M., Jiang H.Q., Huang Y., Qin X.X., Zhao X.T.,
RA Wan D.P., Gu J.R.;
RT "Novel human cDNA clones with function of inhibiting cancer cell
RT growth";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF318323; AAL55830.1; .
DR InterPro: IPR00182; ASX hydroxyl.
DR InterPro: IPR00181; EGF Ca.
DR PROSITE: PS0010; ASX HYDROXYL; 1.
DR PROSITE: PS01187; EGF Ca; 1.
DR Hypothetical protein: EGF-like domain.
SQ SEQUENCE 310 AA; 32607 NW; 7BE18FEDEACBA3B CRC64;
Query Match 9.1%; Score 73; DB 4; Length 310;
Best Local Similarity 24.0%; Pred. No. 5.4;
Matches 31; Conservative 11; Mismatches 45; Indels 42; Gaps 7;
Y 12 SETHAM-----CWSSSTTRSPSRHRLHRIPLALGVTAICSLVMHSHGGISVAL 65
b 202 SSPHSALLSWSCMCPGPPGFLASH-SRPFGLAALQLLCP-----GGRKSSQ 251
Y 66 CSQCLOTNALRP-----DCLTNGGCGYCHGSLGHVDRFPQHSEW 109
Y 252 AS--TSPGASQPCFCTDPASALICSPADKDCSKDNGCQQDC-----VNTFGSYECQC 303
110 NSGMDCKP 118
304 RSGF--CPP 310
ULT 15
832
O08832 PRELIMINARY; PRT: 578 AA.
01-JUL-1997 (TRENBLrel. 04, Created)
01-JUL-1997 (TRENBLrel. 04, Last sequence update)
01-MAR-2003 (TRENBLrel. 23, Last annotation update)
Polypeptide N-acetyl-galactosaminyltransferase (EC 2.4.1.41) (Protein-
UDP acetyl-galactosaminyltransferase) (UDP-GalNAc:polypeptide, N-
acetyl-galactosaminyltransferase) (GalNAc-T4)
GALNT4 OR PEGANTASE-T4
Mus musculus (Mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
(1)
SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
EDLINE=97298094; PubMed=9153242;
agen F.K., Ten Hagen K.G., Beres T.M., Balys M.M.,
anuyckhuysse B.C., Tabak L.A.;
cDNA cloning and expression of a novel UDP-N-acetyl-D-
lactosamine:polypeptide N-Acetyl-galactosaminyltransferase.";
Biol. Chem. 272:13843-13848(1997).
SEQUENCE FROM N.A.
RAIN=c578L/6J; TISSUE=Colon;
EDLINE=22354683; PubMed=12456851;
FANTOM Consortium,
analysis of the mouse transcriptome based on functional annotation of
770 full-length cDNAs.";
ure 420:563-573(2002).
FUNCTION: THIS PROTEIN CATALYZES THE INITIAL REACTION IN O-LINKED
OLIGOSACCHARIDE BIOSYNTHESIS. THE TRANSFER OF AN N-ACETYL-D-
GALACTOSAMINE RESIDUE TO A SERINE OR THREONINE RESIDUE ON THE
PROTEIN RECEPTOR. AGT AND PAG4 HAVE DIFFERENT SUBSTRATES

CC SPECIFICITY.
CC -1- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GALACTOSAMINE + POLYPEPTIDE -
CC UDP + N-ACETYL-D-GALACTOSAMINYL-POLYPEPTIDE.
CC -1- PATHWAY: GLYCOSYLATION.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.
CC -1- TISSUE SPECIFICITY: HIGHER EXPRESSION IN SUBLINGUAL GLAND,
CC STOMACH, SMALL INTESTINE, CERVIX AND COLICUS TO A LESSER EXTENT
CC IN KIDNEY, OVARY, LUNG AND UTERUS. LOW LEVELS IN SPLEEN AND TRACE
CC LEVELS IN LIVER, HEART AND BRAIN. NO EXPRESSION IN SUBMANDIBULAR
CC AND PAROTID GLANDS, SKELETAL MUSCLE AND TESTIS.
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
DR EMBL: U73819; AAB58301.1; .
DR EMBL: AK033494; BAC28317.1; .
DR MGD; MG1:894682; Gainc4.
DR InterPro: IPR001173; Glyco trans 2.
DR InterPro: IPR000772; Ricin_B_lectin.
DR Pfam: PF00835; Glycosyl transfer 2; 1.
DR Pfam: PF00852; Ricin_B_lectin; 2.
DR SMART: SM00458; RICIN_B_LECTIN; 1.
DR PROSITE: PS00311; RICIN_B_LECTIN; 1.
DR Transferrase; glycosyltransferase; Transmembrane, Signal-anchor;
KW Golgi stack; glycoprotein.
FT DOMAIN 1 12 CYTOPLASMIC (POTENTIAL)
FT TRANSMEM 13 31 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT DOMAIN 32 578 LUMENAL CATALYTIC (POTENTIAL)
FT CARBOHYD 471 471 N-LINKED GLCNAC. 1 (POTENTIAL)
SQ SEQUENCE 578 AA; 66554 MW; 10ADC0D8B893085 CRC64;
Query Match 9.0%; Score 72.5; DB 11; Length 578;
Best Local Similarity 23.4%; Pred. No. 12; Indels 55; Gaps 9;
Matches 37; Conservative 20; Mismatches 46;
OY 28 GRHLHRIPLALGVTAICSLVM-----IKYS-----HGGISVALCSQCLOTN 73
Db 410 SERKLRLERLCK-----SFDWYLVNVSNLVPEORPGWHGAIRSMGISSECLDYN 461
OY 74 ALAPRDPCLTNGGCGYCHGSLGHVDRFPQHSE---WNSGMDCK--PLRGPPLG--- 125
Db 462 A--PDNPTGANSLSFG--CHGQGN--QPEFTSKNIRFNSVTELCAEVPOQDYVGMQN 517
OY 126 -----VLTPEPHEPAATRAK 142
Db 518 CPKDGLPVPIVNIHPKEDGTIFPHITRLCLSNRTAE 555
Search completed: November 13, 2003, 13:20:57
Job time : 37 sec

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 13:18:48 ; Search time 21 Seconds
(without alignments)
288.117 Million cell updates/sec

Title: US-09-868-025-2

Perfect score: 803

Sequence: 1 MVDNFPISQSETHAWCMS.....LGVLTTPKMEFAAIRAGKV 143

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents_AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	74	9.2	550	4	US-09-252-991A-27974
2	73.5	9.2	594	4	US-09-252-991A-26272
3	72	9.0	137	4	US-09-252-991A-30560
4	71	8.8	96	4	US-09-252-991A-27861
5	67.5	8.4	431	4	US-09-252-991A-25197
6	67	8.3	339	1	US-08-208-007A-11
7	67	8.3	339	1	US-08-330-121B-7
8	67	8.3	339	3	US-08-860-255A-7
9	67	8.3	339	4	US-08-915-095A-11
10	67	8.3	339	4	US-08-798-096-11
11	67	8.3	339	4	US-08-798-095A-11
12	67	8.3	339	4	US-09-953-956-11
13	67	8.3	339	4	US-08-553-135A-11
14	67	8.3	339	4	US-08-536-881-7
15	67	8.3	339	5	PCT-US95-13820-7
16	66.5	8.3	652	2	US-08-751-305-2
17	66.5	8.3	2259	3	US-09-413-814-70
18	66.5	8.3	2439	3	US-09-335-409-7
19	66.5	8.3	2439	4	US-09-568-102-7
20	66.5	8.3	2439	4	US-09-567-965-7
21	66.5	8.3	2439	4	US-09-568-480-7
22	66.5	8.3	2439	4	US-09-568-486-7
23	66.5	8.3	2439	4	US-09-568-472-7
24	66.5	8.3	2439	4	US-09-567-899-7
25	66	8.2	247	1	US-08-324-977-44
26	66	8.2	247	2	US-08-384-616-44
27	66	8.2	247	2	US-08-904-688A-44

Sequence 44, Appl
Sequence 161, App
Sequence 11, Appl
Sequence 11, Appl
Sequence 11, Appl
Sequence 160, App
Sequence 9, Appl
Sequence 12, Appl
Sequence 12, Appl
Sequence 4, Appl
Sequence 11, Appl
Sequence 18, Appl
Sequence 12, Appl
Sequence 12, Appl
Sequence 12, Appl
Sequence 12, Appl
Sequence 2, Appl
Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-09-252-991A-27974
; Sequence 27974, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27974
; LENGTH: 550
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27974

Query Match 9.2%; Score 74; DB 4; Length 550;
Best Local Similarity 26.2%; Pred. No. 14;
Matches 28; Conservative 19; Mismatches 34; Indels 26; Gaps 5;
Qy 25 RSPSRHH-----LHRRIPCLALGWTATCISLVHSHGGISVALCSQCLQTNALRPD 80
Db 220 KTRPHRGVLRHR-----RGSVSLAGIVEVEVAVGRLDVALGAAAVETLAL----E 270
Qy 81 CLTNNGCYGECGHSIGHYD-----RFPQHSNWSNMGMSCKPLRGE 122
Db 271 LVQGHAAVLGLLHGIGIGLDLALARP-----SLDQVEDVROQ 309

RESULT 2
US-09-252-991A-26272
; Sequence 26272, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 26272
LENGTH: 594
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26272

Query Match 9.2% Score 73.5; DB 4; Length 594;
Best Local Similarity 28.9%; Pred. No. 17;
Matches 22; Conservative 8; Mismatches 27; Indels 19; Gaps 4;

QY 71 QTNALPRPCLTNNGCYCHGSLGHVDRFPQHSNEWNSGMDSC---KPLRGEFLG--- 125
DB 2 KSNLANPRPSCL---GGRFGR-----TGHEQPWYMGCGCIRGKPMQTQETVGA 47

QY 126 VLTPEPKWEFAAIRAG 141
DB 48 ALGHPNVLAIVAVG 63

RESULT 3
US-09-252-991A-30560
Sequence 30560, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252.991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 30560
LENGTH: 137
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30560

Query Match 9.0% Score 72; DB 4; Length 137;
Best Local Similarity 31.9%; Pred. No. 4.5;
Matches 22; Conservative 5; Mismatches 34; Indels 8; Gaps 2;

QY 16 AWCSSSTTRSPRHRLHRIPLCALGVTAICSLVHSHGGSVALCSQC 69
DB 69 AWRPCGIARHPFRHSFLRPEFP--ASGATSRCCFLWRTSEGGLEGPRLYQLQLCLGS 126

QY 70 LQTNALRPR 78
DB 127 SOTHNLOCR 135

RESULT 4
US-09-252-991A-27861
Sequence 27861, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252.991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 27861
LENGTH: 96

TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27861

Query Match 8.4% Score 71; DB 4; Length 96;
Best Local Similarity 30.0%; Pred. No. 3.7;
Matches 24; Conservative 7; Mismatches 35; Indels 14; Gaps 4;

QY 17 WCVSSSTTRSPRHRLHRIPLCALGVTAICSLVHSHGGSVALCS-QCLOTNAL 75
DB 26 WCVASSMCSTASRWY-----GC-----AVAPSSSPMPSATRMPATSVATSPACTRPSCS 75

QY 76 RRPDCL---TNNGCYCEC 92
DB 76 RYRPRCTSAWISPGRCSSRC 95

RESULT 5
US-09-252-991A-25197
Sequence 25197, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252.991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 25197
LENGTH: 431
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25197

Query Match 8.4% Score 67.5; DB 4; Length 431;
Best Local Similarity 28.9%; Pred. No. 49;
Matches 22; Conservative 12; Mismatches 19; Indels 23; Gaps 3;

QY 28 SRHLHRIPLCALGVTAICSLVHSHGGSVALCSQCLOTNALRPRDCLTNNCG 87
DB 230 ARHVORER---TAVGLDAVTVFRFLAVAHGHLRVG-----EHHGR 268

QY 88 CYGECHGSL--GHVDR 101
DB 269 HGGQGVGSVATGHVDR 284

RESULT 6
US-08-208-007A-11
Sequence 11, Application US/08208007A
Patent No. 5501969
GENERAL INFORMATION:
APPLICANT: HASTINGS, ET AL.
TITLE OF INVENTION: Human Osteoclast-Derived Cathepsin
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
ADDRESS: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1

NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 26272
LENGTH: 594
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26272

Query Match 9.2% Score 73.5; DB 4; Length 594;
Best Local Similarity 28.9%; Pred. No. 17;
Matches 22; Conservative 8; Mismatches 27; Indels 19; Gaps 4;

QY 71 QTNALPRPCLTNNGCYCHGSLGHVDRFPQHSNEWNSGMDSC---KPLRGEFLG--- 125
DB 2 KSNLANPRPSCL---GGRFGR-----TGHEQPWYMGCGCIRGKPMQTQETVGA 47

QY 126 VLTPEPKWEFAAIRAG 141
DB 48 ALGHPNVLAIVAVG 63

RESULT 3
US-09-252-991A-30560
Sequence 30560, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252.991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 30560
LENGTH: 137
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30560

Query Match 9.0% Score 72; DB 4; Length 137;
Best Local Similarity 31.9%; Pred. No. 4.5;
Matches 22; Conservative 5; Mismatches 34; Indels 8; Gaps 2;

QY 16 AWCSSSTTRSPRHRLHRIPLCALGVTAICSLVHSHGGSVALCSQC 69
DB 69 AWRPCGIARHPFRHSFLRPEFP--ASGATSRCCFLWRTSEGGLEGPRLYQLQLCLGS 126

QY 70 LQTNALRPR 78
DB 127 SOTHNLOCR 135

RESULT 4
US-09-252-991A-27861
Sequence 27861, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252.991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 27861
LENGTH: 96

TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27861

Query Match 8.4% Score 71; DB 4; Length 96;
Best Local Similarity 30.0%; Pred. No. 3.7;
Matches 24; Conservative 7; Mismatches 35; Indels 14; Gaps 4;

QY 17 WCVSSSTTRSPRHRLHRIPLCALGVTAICSLVHSHGGSVALCS-QCLOTNAL 75
DB 26 WCVASSMCSTASRWY-----GC-----AVAPSSSPMPSATRMPATSVATSPACTRPSCS 75

QY 76 RRPDCL---TNNGCYCEC 92
DB 76 RYRPRCTSAWISPGRCSSRC 95

RESULT 5
US-09-252-991A-25197
Sequence 25197, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252.991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 25197
LENGTH: 431
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25197

Query Match 8.4% Score 67.5; DB 4; Length 431;
Best Local Similarity 28.9%; Pred. No. 49;
Matches 22; Conservative 12; Mismatches 19; Indels 23; Gaps 3;

QY 28 SRHLHRIPLCALGVTAICSLVHSHGGSVALCSQCLOTNALRPRDCLTNNCG 87
DB 230 ARHVORER---TAVGLDAVTVFRFLAVAHGHLRVG-----EHHGR 268

QY 88 CYGECHGSL--GHVDR 101
DB 269 HGGQGVGSVATGHVDR 284

RESULT 6
US-08-208-007A-11
Sequence 11, Application US/08208007A
Patent No. 5501969
GENERAL INFORMATION:
APPLICANT: HASTINGS, ET AL.
TITLE OF INVENTION: Human Osteoclast-Derived Cathepsin
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
ADDRESS: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/208,007A
FILING DATE: March 8, 1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: No. 5501969e
FILING DATE: No. 5501969e
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-95
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 339 AMINO ACIDS
TYPE: AMINO ACID
STRANDEDNESS:
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
US-08-208-007A-11

Query Match 8.3%; Score 67; DB 1; Length 339;
Best Local Similarity 22.3%; Pred. No. 42;
Matches 33; Conservative 13; Mismatches 56; Indels 46; Gaps 7;

QY 13 ETHAWCWSSTTRSPSRHHLRERIPCLALGVTAICSLV-----WIHVSHGG--GIS 62
DB 174 ESHVGCPRYSI--PPCEHHVNGSRPCTGEGDTPKCSKICEPGYSTYKQDKHYGNSYS 231
QY 63 VALCSQCLQTNALRPRP-----DCLTNGGCYCECHGSL--GHVDR----- 101
DB 232 VSNSEKIDMAEIKYNGPVEGAFSVYDFLLYKSGVYOHVTGEMMGHGAIRILGWGVNGT 291
QY 102 -FPQHSNEWS-----GMDSC 116
DB 292 PYMLVANSWNTDWDGNGFFKILGGQDHC 319

Query Match 8.3%; Score 67; DB 1; Length 339;
Best Local Similarity 22.3%; Pred. No. 42;
Matches 33; Conservative 13; Mismatches 56; Indels 46; Gaps 7;

QY 13 ETHAWCWSSTTRSPSRHHLRERIPCLALGVTAICSLV-----WIHVSHGG--GIS 62
DB 174 ESHVGCPRYSI--PPCEHHVNGSRPCTGEGDTPKCSKICEPGYSTYKQDKHYGNSYS 231
QY 63 VALCSQCLQTNALRPRP-----DCLTNGGCYCECHGSL--GHVDR----- 101
DB 232 VSNSEKIDMAEIKYNGPVEGAFSVYDFLLYKSGVYOHVTGEMMGHGAIRILGWGVNGT 291
QY 102 -FPQHSNEWS-----GMDSC 116
DB 292 PYMLVANSWNTDWDGNGFFKILGGQDHC 319

US-08-330-121B-7
Sequence 7, Application US/08330121B
Patent No. 5736357
GENERAL INFORMATION:
APPLICANT: Bromme, Dieter
APPLICANT: Okamoto, Kathleen
TITLE OF INVENTION: CATHEPSIN O2 PROTEASE
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hobbach, Test, Albritton
ADDRESSEE: & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/330,121B
FILING DATE: 27-OCT-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-60261/DJB/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249

TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 339 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-330-121B-7

Query Match 8.3%; Score 67; DB 1; Length 339;
Best Local Similarity 22.3%; Pred. No. 42;
Matches 33; Conservative 13; Mismatches 56; Indels 46; Gaps 7;

QY 13 ETHAWCWSSTTRSPSRHHLRERIPCLALGVTAICSLV-----WIHVSHGG--GIS 62
DB 174 ESHVGCPRYSI--PPCEHHVNGSRPCTGEGDTPKCSKICEPGYSTYKQDKHYGNSYS 231
QY 63 VALCSQCLQTNALRPRP-----DCLTNGGCYCECHGSL--GHVDR----- 101
DB 232 VSNSEKIDMAEIKYNGPVEGAFSVYDFLLYKSGVYOHVTGEMMGHGAIRILGWGVNGT 291
QY 102 -FPQHSNEWS-----GMDSC 116
DB 292 PYMLVANSWNTDWDGNGFFKILGGQDHC 319

RESULT 8
US-08-860-255A-7
Sequence 7, Application US/08860255A
Patent No. 6274336
GENERAL INFORMATION:
APPLICANT: Abdel-Meqid, Sherin
APPLICANT: Desjarlais, Renee
APPLICANT: Janson, Cheryl
APPLICANT: Smith, Ward
APPLICANT: Zhao, Baoquang
TITLE OF INVENTION: Method of Inhibiting Cathepsin K
FILE REFERENCE: P50574-XI
CURRENT APPLICATION NUMBER: US/08/860,255A
CURRENT FILING DATE: 1997-06-26
PRIOR APPLICATION NUMBER: 60/008,108
PRIOR FILING DATE: 1995-10-30
PRIOR APPLICATION NUMBER: 60/007,473
PRIOR FILING DATE: 1995-11-22
PRIOR APPLICATION NUMBER: 60/008,592
PRIOR FILING DATE: 1995-12-21
PRIOR APPLICATION NUMBER: 60/013,748
PRIOR FILING DATE: 1996-03-20
PRIOR APPLICATION NUMBER: 60/013,764
PRIOR FILING DATE: 1996-03-20
PRIOR APPLICATION NUMBER: 60/013,747
PRIOR FILING DATE: 1996-03-20
PRIOR APPLICATION NUMBER: 60/017,455
PRIOR FILING DATE: 1996-05-17
PRIOR APPLICATION NUMBER: 60/017,892
PRIOR FILING DATE: 1996-05-17
PRIOR APPLICATION NUMBER: 60/020,478
PRIOR FILING DATE: 1996-06-13
PRIOR APPLICATION NUMBER: 60/022,047
PRIOR FILING DATE: 1996-07-22
PRIOR APPLICATION NUMBER: 60/023,494
PRIOR FILING DATE: 1996-08-07
PRIOR APPLICATION NUMBER: 60/023,742
PRIOR FILING DATE: 1996-08-08
NUMBER OF SEQ ID NOS: 7
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO: 7
LENGTH: 339
TYPE: PRT
ORGANISM: homo sapiens
US-08-860-255A-7

Query Match 8.3%; Score 67; DB 3; Length 339;
Best Local Similarity 22.3%; Pred. No. 42;
Matches 33; Conservative 13; Mismatches 56; Indels 46; Gaps 7;

QY 13 ETHAWCWSSTTRSPRHLHRRIPCLALGVTAICSLV-----WIHVSHGG--GIS 62
DB 174 ESHVGCRRPYSI--PPCEHHVNGSRPPCTGEGDTPKCKICPCPGYSPTYKQDKHYGYSYS 231
QY 63 VALCSQCLQTNALRPRP-----DCLTNNGGCYGECHGSL--GHVDR----- 101
DB 232 VSNSEKDIMAEIYKNGPVEGAFSVYSDFLLYKSGVYQHVVTGEMMGHRAIRILGWGVNGT 291
QY 102 -FPQHSNEWS-----GMDSC 116
DB 292 PYWLANSWNTDWDGNGFFKILRGQDHC 319

RESULT 9

US-08-915-095A-11
; Sequence 11, Application US/08915095A

; Patent No. 6387933

; GENERAL INFORMATION:

; APPLICANT: Hastings, et al.

; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN

; FILE REFERENCE: PF107D4

; CURRENT APPLICATION NUMBER: US/08/915,095A

; CURRENT FILING DATE: 1997-08-20

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 11

; LENGTH: 339

; TYPE: PRT

; ORGANISM: Homo sapiens

US-08-915-095A-11

Query Match 8.3%; Score 67; DB 4; Length 339;
Best Local Similarity 22.3%; Pred. No. 42;
Matches 33; Conservative 13; Mismatches 56; Indels 46; Gaps 7;

QY 13 ETHAWCWSSTTRSPRHLHRRIPCLALGVTAICSLV-----WIHVSHGG--GIS 62
DB 174 ESHVGCRRPYSI--PPCEHHVNGSRPPCTGEGDTPKCKICPCPGYSPTYKQDKHYGYSYS 231
QY 63 VALCSQCLQTNALRPRP-----DCLTNNGGCYGECHGSL--GHVDR----- 101
DB 232 VSNSEKDIMAEIYKNGPVEGAFSVYSDFLLYKSGVYQHVVTGEMMGHRAIRILGWGVNGT 291
QY 102 -FPQHSNEWS-----GMDSC 116
DB 292 PYWLANSWNTDWDGNGFFKILRGQDHC 319

RESULT 10

US-08-798-096-11
; Sequence 11, Application US/08798096

; Patent No. 6387682

; GENERAL INFORMATION:

; APPLICANT: Hastings, et al.

; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN

; FILE REFERENCE: PF107D2

; CURRENT APPLICATION NUMBER: US/08/798,096

; CURRENT FILING DATE: 1997-02-12

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 11

; LENGTH: 339

; TYPE: PRT

; ORGANISM: Homo sapiens

US-08-798-096-11

Query Match 8.3%; Score 67; DB 4; Length 339;
Best Local Similarity 22.3%; Pred. No. 42;
Matches 33; Conservative 13; Mismatches 56; Indels 46; Gaps 7;

QY 13 ETHAWCWSSTTRSPRHLHRRIPCLALGVTAICSLV-----WIHVSHGG--GIS 62
DB 174 ESHVGCRRPYSI--PPCEHHVNGSRPPCTGEGDTPKCKICPCPGYSPTYKQDKHYGYSYS 231
QY 63 VALCSQCLQTNALRPRP-----DCLTNNGGCYGECHGSL--GHVDR----- 101
DB 232 VSNSEKDIMAEIYKNGPVEGAFSVYSDFLLYKSGVYQHVVTGEMMGHRAIRILGWGVNGT 291
QY 102 -FPQHSNEWS-----GMDSC 116
DB 292 PYWLANSWNTDWDGNGFFKILRGQDHC 319

RESULT 11

US-08-798-095A-11
; Sequence 11, Application US/08798095A

; Patent No. 6423507

; GENERAL INFORMATION:

; APPLICANT: Hastings, et al.

; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN

; FILE REFERENCE: PF107D3

; CURRENT APPLICATION NUMBER: US/08/798,095A

; CURRENT FILING DATE: 1997-02-12

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 11

; LENGTH: 339

; TYPE: PRT

; ORGANISM: Homo sapiens

US-08-798-095A-11

Query Match 8.3%; Score 67; DB 4; Length 339;
Best Local Similarity 22.3%; Pred. No. 42;
Matches 33; Conservative 13; Mismatches 56; Indels 46; Gaps 7;

QY 13 ETHAWCWSSTTRSPRHLHRRIPCLALGVTAICSLV-----WIHVSHGG--GIS 62
DB 174 ESHVGCRRPYSI--PPCEHHVNGSRPPCTGEGDTPKCKICPCPGYSPTYKQDKHYGYSYS 231
QY 63 VALCSQCLQTNALRPRP-----DCLTNNGGCYGECHGSL--GHVDR----- 101
DB 232 VSNSEKDIMAEIYKNGPVEGAFSVYSDFLLYKSGVYQHVVTGEMMGHRAIRILGWGVNGT 291
QY 102 -FPQHSNEWS-----GMDSC 116
DB 292 PYWLANSWNTDWDGNGFFKILRGQDHC 319

RESULT 12

US-09-953-956-11
; Sequence 11, Application US/09953956

; Patent No. 6475487

; GENERAL INFORMATION:

; APPLICANT: Hastings, et al.

; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN

; FILE REFERENCE: PF107D2D1

; CURRENT APPLICATION NUMBER: US/09/953,956

; CURRENT FILING DATE: 2001-09-18

; PRIOR APPLICATION NUMBER: 09/219,441

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 11

; LENGTH: 339

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-953-956-11

Query Match 8.3%; Score 67; DB 4; Length 339;
Best Local Similarity 22.3%; Pred. No. 42;
Matches 33; Conservative 13; Mismatches 56; Indels 46; Gaps 7;

QY 13 ETHAWCWSSTTRSPSRHHLRERIPCLALGVTAICSLV-----WIHVSHGG--GIS 62
DB 174 ESHVGCPRYSI--PPCEHVNGSRPCTGEGDTPKCKICPCGYSTYKQDKHYGYSYS 231
QY 63 VALCSQCLQTNALRPP-----DCLTNNGGCYGECHGSL--GHVDR-----101
DB 232 VSNSEKDIMAEIYKNGPVEGAFSVYDFLLYKSGVYQHVVTGEMMGHAIIRILGWGVNGT 291
QY 102 -FPQHSNEWNS-----GMDSC 116
DB 292 PYWLVSANWTDGNDGNGFFKILRGQDHC 319

RESULT 13
US-08-553-125A-11
; Sequence 11, Application US/08553125A
; Patent No. 6475766
; GENERAL INFORMATION:
; APPLICANT: Hastings, et al.
; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
; FILE REFERENCE: PFI07D1
; CURRENT APPLICATION NUMBER: US/08/553,125A
; CURRENT FILING DATE: 1995-11-07
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 11
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-553-125A-11

Query Match 8.3%; Score 67; DB 4; Length 339;
Best Local Similarity 22.3%; Pred. No. 42;
Matches 33; Conservative 13; Mismatches 56; Indels 46; Gaps 7;

QY 13 ETHAWCWSSTTRSPSRHHLRERIPCLALGVTAICSLV-----WIHVSHGG--GIS 62
DB 174 ESHVGCPRYSI--PPCEHVNGSRPCTGEGDTPKCKICPCGYSTYKQDKHYGYSYS 231
QY 63 VALCSQCLQTNALRPP-----DCLTNNGGCYGECHGSL--GHVDR-----101
DB 232 VSNSEKDIMAEIYKNGPVEGAFSVYDFLLYKSGVYQHVVTGEMMGHAIIRILGWGVNGT 291
QY 102 -FPQHSNEWNS-----GMDSC 116
DB 292 PYWLVSANWTDGNDGNGFFKILRGQDHC 319

RESULT 14
US-08-536-861-7
; Sequence 7, Application US/08536861
; Patent No. 6544767
; GENERAL INFORMATION:
; APPLICANT: Bromme, Dieter
; APPLICANT: Okamoto, Kathleen
; TITLE OF INVENTION: CATHEPSIN O2 PROTEASE
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohnbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/536,861
; FILING DATE: 02-OCT-1995
; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-60261-1/DJB/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 339 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-536-861-7

Query Match 8.3%; Score 67; DB 4; Length 339;
Best Local Similarity 22.3%; Pred. No. 42;
Matches 33; Conservative 13; Mismatches 56; Indels 46; Gaps 7;

QY 13 ETHAWCWSSTTRSPSRHHLRERIPCLALGVTAICSLV-----WIHVSHGG--GIS 62
DB 174 ESHVGCPRYSI--PPCEHVNGSRPCTGEGDTPKCKICPCGYSTYKQDKHYGYSYS 231
QY 63 VALCSQCLQTNALRPP-----DCLTNNGGCYGECHGSL--GHVDR-----101
DB 232 VSNSEKDIMAEIYKNGPVEGAFSVYDFLLYKSGVYQHVVTGEMMGHAIIRILGWGVNGT 291
QY 102 -FPQHSNEWNS-----GMDSC 116
DB 292 PYWLVSANWTDGNDGNGFFKILRGQDHC 319

RESULT 15
PCT-US95-13820-7
; Sequence 7, Application PC/TUS9513820
; GENERAL INFORMATION:
; APPLICANT: Khepri Pharmaceuticals, Inc.
; TITLE OF INVENTION: CATHEPSIN O2 PROTEASE
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohnbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13820
; FILING DATE: 26-OCT-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US UNKNOWN
; FILING DATE: 02-OCT-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/330,121
; FILING DATE: 27-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: PP-60261-1-PC/DJB/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:

Search completed: November 13, 2003, 13:22:03
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 13:18:42 ; Search time 41 Seconds

(without alignments)
553.607 Million cell updates/sec

Title: US-09-868-025-2

Perfect score: 803

Sequence: 1 MVDNFIQSETHNCWS.....LGVLTTPKMEFAAIRAGKV 143

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_13Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	803	100.0	143	22	Rice AGT-SAL 11 pr
2	86.5	11.0	274	23	Human polypeptide
3	81.5	10.1	262	22	Human gene 8 encod
4	78	9.7	257	20	Chicken 17.5.3 pro
5	74.5	9.3	9222	22	Novel human diagno
6	74	9.2	4561	22	Novel human diagno
7	73.5	9.2	478	22	Novel human diagno
8	73	9.1	139	22	Novel human secret
9	73	9.1	320	13	HCV antigen peptid

10	73	9.1	1463	20	AAW99482	Murine NCoA-2 prot
11	71	8.8	131	20	AAV19454	Amino acid sequenc
12	71	8.8	286	22	AAU31738	Novel human secret
13	71	8.8	1686	22	AAZ00913	Human 27875 ADAM-T
14	71	8.8	1686	22	AAZ00914	Human 27875 ADAM-T
15	70.5	8.8	143	22	ABG26623	Novel human diagno
16	70.5	8.8	188	22	AAE11760	Mouse mast cell fu
17	70.5	8.8	874	23	ABP43970	FLJ10300 fis clone
18	70.5	8.8	1679	22	ABP40498	Drosophila melanog
19	70.5	8.8	1679	22	ABP40498	Drosophila melanog
20	70	8.7	222	13	AAZ25885	HK10. Hepatitis C
21	70	8.7	1037	22	AAU07142	Mouse CRIM1 protei
22	70	8.7	1188	13	AAZ29870	HCV NS2-NS4 peptid
23	69.5	8.7	323	20	AAV74088	Human prostate tum
24	69	8.6	243	21	AAE38477	Fragment of human
25	69	8.6	1035	23	ABB98128	Human PMW incyte
26	69	8.6	1094	23	AAU72900	Human metalloprote
27	69	8.6	1431	23	AAW47675	Rat Hyaluronidase
28	69	8.6	1431	24	ABG72498	Rat 175kDa Hyaluro
29	69	8.6	1464	22	ABB71111	Drosophila melanog
30	68.5	8.5	110	22	AAU42751	Propionibacterium
31	68.5	8.5	1036	21	AAZ82776	Human chordin rela
32	68.5	8.5	1036	21	AAV53034	Human secreted pro
33	68.5	8.5	1036	22	AAU07141	Human CRIM1 protei
34	68.5	8.5	1036	22	AAU12242	Human PRO4330 poly
35	68.5	8.5	1036	23	AAE18852	Human pharmaceutical
36	68.5	8.5	1036	24	ABU66640	Human PRO polypept
37	68.5	8.5	1036	24	ABU66616	Human secreted/tra
38	68.5	8.5	1036	24	ABU59721	Novel secreted and
39	68	8.5	125	22	AAU40360	Propionibacterium
40	68	8.5	194	15	AAZ52737	HCV antigen. Synt
41	68	8.5	195	13	AAZ25863	HCV polypeptide 10
42	68	8.5	195	14	AAW41741	Hepatitis C virus
43	68	8.5	293	17	AAZ98350	HCV NS4 derived an
44	68	8.5	388	20	AAV04998	Mycobacterium spec
45	68	8.5	477	13	AAZ29865	HCV NS2-NS4 peptid

ALIGNMENTS

RESULT 1
ID AAB75128 standard; Protein; 143 AA.
AC AAB75128;
DT 03-AUG-2001 (first entry)
DE Rice AGT-SAL 11 protein sequence SEQ ID NO:2.
KW Rice AGT-SAL 11; salt tolerance; plant; cytostatic; antiHIV;
KW proteinase inhibitor; cancer; human immunodeficiency virus;
KW HIV infection; animal disorder; food processing; enzyme industry;
KW biological preservative.
OS Oryza sativa.
XX MO200130990-A2.
XX 03-MAY-2001.
XX 11-OCT-2000; 2000MO-IN00099.
XX 13-OCT-1999; 99IN-0000997.
XX (AVES-) AVESTHAGEN GRAINE TECHNOLOGIES PVT LTD.
XX Patell VM, Antony CM, Chandran D, Madurappa A;
XX WPI; 2001-308632/32.
XX N-FSD; AAH19867.

PT Nucleotide sequence encoding an AGT-SAL 11 polypeptide similar to
PT Bowman Birk II type proteinase inhibitors is useful to confer salt
XX resistance to plants
XX
PS Claim 2; Page 17; 22pp; English.
XX
CC The present sequence represents the rice AGT-SAL 11 protein. The
CC AGT-SAL 11 protein can be used to confer salt tolerance to plants and
CC other organisms. The AGT-SAL 11 gene was isolated from salt-stressed
CC rice. Also described are: (i) a transgenic plant comprising a recombinant
CC expression cassette comprising a plant promoter operably linked to N1;
CC and (ii) conferring salt tolerance on a plant, comprising introducing
CC the above expression cassette. The AGT-SAL 11 protein is a proteinase
CC inhibitor. The proteinase inhibitor may be used to confer stress
CC tolerance to many plants and organisms including cotton, maize, rice,
CC soybean, sugar beet, wheat, fruit, vegetables and vines, particularly
CC biotic bacterial, fungal and pest stresses. Proteinase inhibitors are
CC also useful in the treatment of cancer, human immunodeficiency virus
CC (HIV) infection and other animal disorders. The gene may also be useful
CC in food processing and enzyme industries as an inhibitor of proteinase
CC activity and as a biological preservative.
XX
SQ Sequence 143 AA;
Query Match 100.0%; Score 803; DB 22; Length 143;
Best Local Similarity 100.0%; Pred. No. 5e-75;
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDNFPISSEQSETHAWCWSSTTRSPRHLHRRIPCLALGVTAICSLVWVHSHGGG 60
DB 1 MVDNFPISSEQSETHAWCWSSTTRSPRHLHRRIPCLALGVTAICSLVWVHSHGGG 60
QY 61 ISVALCSQCLQTNALRPDCLTNNGGCYGCHGSLGHVDRFPQHSNENSGMDCKPLR 120
DB 61 ISVALCSQCLQTNALRPDCLTNNGGCYGCHGSLGHVDRFPQHSNENSGMDCKPLR 120
QY 121 GEFGLVLTTPHPMEPAARAGKV 143
DB 121 GEFGLVLTTPHPMEPAARAGKV 143
RESULT 2
ABP69692 ID ABP69692 standard; Protein; 274 AA.
XX AC ABP69692;
XX DT 20-JAN-2003 (first entry)
XX DE Human polypeptide SEQ ID NO 1739.
XX KW Human; genome mapping; gene therapy; food supplement; virus; fungus;
KW cell-proliferative disorder; neurodegenerative disease; bacterial;
KW Parkinson's disease; Alzheimer's disease; autoimmune disease;
KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
KW arthritis; cytostatic; immunomodulator; nontropic; neuroprotective;
KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;
KW haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;
KW antiarthritic.
XX OS Homo sapiens.
XX PN WO200270539-A2.
XX PD 12-SEP-2002.
XX PF 05-MAR-2002; 2002WO-US05095.
XX PR 05-MAR-2001; 2001US-0799451.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;

PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;
PI Wehman T, Wang J, Wang D, Drmanac RT;
XX WPI: 2002-759812/82.
DR N-PSDB; AB211909.
XX
PT New polynucleotides comprising sequences assembled from expressed
PT sequence tags (ESTs), useful for treating cell-proliferative,
PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or
PT platelet or coagulation disorders
XX
PS Claim 9; SEQ ID NO 1739; 1012pp + Sequence Listing; English.
XX
CC The invention relates to an isolated polynucleotide (i) comprising a
CC nucleotide sequence selected from any of 948 sequences
CC (AB21119-AB212066) or their mature protein coding portion, active domain
CC coding protein or complementary sequences. The polynucleotides are useful
CC for identifying expressed genes or for physical mapping of human genome.
CC The encoded polypeptides (ABP6902-ABP6949) are useful as molecular
CC weight markers, as a food supplement, for generating antibodies, in
CC medical imaging, screening and diagnostic assays and for treating
CC cell-proliferative disorders (cancer), neurodegenerative diseases
CC (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple
CC sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid
CC disorders, platelet or coagulation disorders, wound, burn, incision,
CC ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,
CC parasitic), arthritis, etc.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 274 AA;
Query Match 11.0%; Score 88.5; DB 23; Length 274;
Best Local Similarity 26.5%; Pred. No. 0.59;
Matches 35; Conservative 14; Mismatches 40; Indels 43; Gaps 8;
QY 13 ETHAWCWSSTTRSPRHLHRRIPCLALGVTAICSLVWVHSHGGISVALCSOC--- 69
DB 158 EPPHAW---TDRPHH-----GGVQRVCPQVPLPRPGGLSPHCSGGLP 203
QY 70 --LQTNALRP---RPDCLTNNGGCYGCH--GSLGHVDRFPQHSNENSGMDCKPLRG 122
DB 204 QHCSHLLKPVSSKCPSLTS-----ECHRPPSQNYLSRV---KNWAG-----GD 244
QY 123 FLOGLTTPHPKME 134
DB 245 FLIVLPQMOLE 256
RESULT 3
AAE04267 ID AAE04267 standard; Protein; 262 AA.
XX AC AAE04267;
XX DT 09-AUG-2001 (first entry)
XX DE Human gene 8 encoded secreted protein fragment, SEQ ID NO:130.
XX KW Human; secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiotensin disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder; gene therapy;
KW endocrine disorder; infection; wound healing; vulnery;
KW cell culture; chemotaxis; food additive;
KW binding partner identification; chromosome 7.
XX OS Homo sapiens.

XX PN WO200136432-A2.
XX PD 25-MAY-2001.
XX PF 15-NOV-2000; 2000WO-US11162.
XX PR 19-NOV-1999; 99US-0166415.
XX PR 30-JUN-2000; 2000US-0215136.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Ruben SM, Komatsoulis GA, Baker KP, Young PB;
XX PD WPI; 2001-343793/36.
XX PT Isolated nucleic acid molecule encoding a human secreted protein is
XX PT used in preventing, treating or ameliorating a medical condition -
XX PS Disclosure; Page 32; 509pp; English.
XX CC AAD08488-AAD08529 represent cDNAs corresponding to 18 human secreted
XX CC protein genes, and AAE04199-AAE04239 represent the proteins they encode.
XX CC AAE04240-AAE04297 represent human secreted protein fragments or variants.
XX CC The secreted proteins and their genes are useful for preventing,
XX CC treating or ameliorating medical conditions, e.g., by protein or gene
XX CC amount of the new protein in a sample or by determining the presence of
XX CC mutations in the new genes. Specific uses are described for each of the
XX CC 18 genes, based on the tissues in which they are most highly expressed,
XX CC and include developing products for the diagnosis or treatment of
XX CC proliferative disorders, cancer, tumours, foetal and developmental
XX CC abnormalities, haematopoietic disorders, diseases of the immune system,
XX CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
XX CC allergies, neurological disorders (e.g., Alzheimer's disease,
XX CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
XX CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
XX CC cardiovascular disorders, angiotensin disorders, kidney disorders,
XX CC gastrointestinal disorders, pregnancy-related disorders, endocrine
XX CC disorders, and infections. The proteins can also be used to aid wound
XX CC healing and epithelial cell proliferation, to prevent skin ageing due to
XX CC sunburn, to maintain organs before transplantation, for supporting cell
XX CC culture of primary tissues, to regenerate tissues, to identify their
XX CC cognate ligands or binding partners, and in chemotaxis, and can be used
XX CC as a food additive or preservative to modify storage properties.
XX CC Antibodies specific for a protein of the invention can be used in
XX CC alleviating symptoms associated with the disorders mentioned above, and
XX CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
XX CC immunoassay (ELISA). The present sequence represents a human
XX CC secreted protein fragment referred to in the disclosure of the invention.
XX SQ Sequence 262 AA;
Query Match 10.1%; Score 81.5; DB 22; Length 262;
Best Local Similarity 27.4%; Pred. No. 2.9;
Matches 26; Conservative 10; Mismatches 38; Indels 21; Gaps 4;
QY 52 WIVSHGGGIVSALCSQ-CLQTNALRPDPD-----LTNNGGCGVGECHGS 95
DB 7 WQVAPG---BILGCEQTCLENNATKTSNCSSARASGCVQCPGHRFSAQPCVPEDHCB 63
QY 96 LGHVDRFPQHSNEWNSGMDCKPLRGFLGLVTPH 130
DB 64 CWHLGRPHLPGEWQEAECSCILSGR--PVCTQH 96
RESULT 4
ID AAW85594 standard; Protein; 257 AA.
XX AAW85594;
XX AC AAW85594;
XX DT 02-MAR-1999 (first entry)

XX DE Chicken 17.5.3 protein.
XX KW lectin; carbohydrate; binding; agglutination; selectin; receptor;
XX KW calcium dependent binding; treatment; cancer; detection;
XX KW identification; CTL-1; ovary; paraganglion; penis; brain; thyroid;
XX KW heart.
XX OS Gallus domesticus.
XX PN WO9849306-A1.
XX PD 05-NOV-1998.
XX PF 29-APR-1998; 98WO-US08791.
XX PR 29-APR-1997; 97US-0846523.
XX PA (INCY-) INCYTE PHARM INC.
XX PI Bandman O, Shah P;
XX DR WPI; 1999-024060/02.
XX DR N-PSDB; AAV83109.
XX PT New human C-type lectin and related nucleic acid, vectors,
XX PT transformed cells - antibodies, agonists and antagonists, for
XX PT diagnosis, prevention and treatment of cancers
XX PS Example 3; Page 48-49; 64pp; English.
XX CC Chicken 17.5.3 protein shows homology with the human CTL-1 protein.
XX CC Recombinant cells containing expression vectors comprising the CTL-1
XX CC coding sequence can be used to produce recombinant CTL-1 which is
XX CC useful for raising Ab and to screen for specific binding agents.
XX CC Binding agents which are antagonists of CTL-1 can be used to
XX CC treat or prevent cancer, e.g. of ovary, paraganglion, penis, brain,
XX CC thyroid and heart which are associated with overexpression of
XX CC CTL-1. Fragments of this sequence are used to diagnose these
XX CC conditions, as probes or primers in usual hybridisation and/or
XX CC amplification assays, or for gene mapping, while complements of
XX CC this sequence, antisense or ribozyme sequences are used to treat or
XX CC prevent the aforementioned cancers also. Ab are used directly as
XX CC antagonists or for delivery of therapeutic agents to cells that
XX CC express CTL-1. CTL-1 includes a carbohydrate-recognition domain
XX CC (aa35-149) and also a transmembrane domain, but no secretory signal.
XX CC CTL-1 also shows homology with human CD69 (AAW85593), and mouse
XX CC CD69 (AAW85595) proteins.
XX SQ Sequence 257 AA;
Query Match 9.7%; Score 78; DB 20; Length 257;
Best Local Similarity 26.9%; Pred. No. 6.6;
Matches 35; Conservative 14; Mismatches 47; Indels 34; Gaps 7;
QY 11 QSETHAWCWSSTTRSPSRHLLHREIPCLALGVTAICSLVWTHVSHGGGIVSALCSQ-C 69
DB 75 QSORGSGC-----SBLQRNRRLVCLVALSAVP-CMLV-----LALVAVIV 113
QY 70 LQTNALRPDPD---CLTNNGGCGVGECHGSLGHVDRFPQHSNEWNSGMDCKPLRGFLG 125
DB 114 LQAPSCSPRPFPFSHVCFNAWVGQKCY-----YFSTESDWSNRCHRL-GASLA 165
QY 126 VLTFHPNMEFP 135
DB 166 TLDTKEMEF 175
RESULT 5
ID ABG21064 standard; Protein; 9222 AA.
XX ABG21064;
XX AC ABG21064;

```

XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #21055.
XX AC
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX PI WPI; 2001-639362/73.
XX DR N-PSDB; AAS85251.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 20; SEQ ID No 51423; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human
XX CC diagnostic amino acid sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 9222 AA;
Query Match 9.3%; Score 74.5; DB 22; Length 9222;
Best Local Similarity 25.2%; Pred. No. 1e+03;
Matches 34; Conservative 11; Mismatches 55; Indels 35; Gaps 6;
QY 15 HAWCNSSTT-----RSPSRHLHRRIPCLALGVTAICSLV-----WIHVSHG 59
DB 3798 HSLPWGLTILTEBQAQELPPGTGLTRNCTRCVCHGAFSCSLVDCQVPPQETWQVAPG- 3856
QY 60 GISVALCSQ-CLQTNALRPPDC-----LTNNGCYGECGSLGHVDRFP 103
DB 3857 --ELGLCQTCLFNATQTQNCSSARAGSCVCGQGHFRSQAGPCVFEDNCECWLGR-P 3913
QY 104 QHSNENWNGMDSCKP 118
DB 3914 HLGEEWVLEPGSCCP 3928

```

```

RESULT 6
ABG30203
ID ABG30203 standard; Protein; 4561 AA.
XX AC
XX ABG30203;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #30194.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX PI WPI; 2001-639362/73.
XX DR N-PSDB; AAS94390.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 20; SEQ ID No 60562; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human
XX CC diagnostic amino acid sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 4561 AA;
Query Match 9.2%; Score 74; DB 22; Length 4561;
Best Local Similarity 25.0%; Pred. No. 5e+02;
Matches 34; Conservative 11; Mismatches 55; Indels 36; Gaps 6;
QY 15 HAWCNSSTT-----RSPSRHLHRRIPCLALGVTAICSLV-----WIHVSHG 58
DB 3074 HSLPWGLTILTEBQAQELPPGTGLTRNCTRCVCHGAFSCSLVDCQVPPQETWQVAPG 3133
QY 59 GISVALCSQ-CLQTNALRPPDC-----LTNNGCYGECGSLGHVDRF 102

```

Db 3134 ---ELGLCEQTCLENNATKTSNCCSARASGVCQPGHFRSQAGPCVPEDHCECHWILGR- 3189

Qy 103 POHSHENWGMDSCKP 118

Db 3190 PHLGEWVLEPGSCCP 3205

RESULT 7

ABG27155

ID ABG27155 standard; Protein; 478 AA.

XX AC

XX DT

XX DT 18-FEB-2002 (first entry)

XX DE

XX DE Novel human diagnostic protein #27146.

XX KW

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.

XX OS

XX OS Homo sapiens.

XX PN

XX PN WO200175067-A2.

XX PD

XX PD 11-OCT-2001.

XX PF

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR

XX PR 31-MAR-2000; 2000US-0540217.

XX PR

XX PR 23-AUG-2000; 2000US-0649167.

XX PA

XX PA (HYSE-) HYSEQ INC.

XX PI

XX PI Drmanac RT, Liu C, Tang YT;

XX DR

XX DR WPI; 2001-639362/73.

XX DR

XX DR N-PSDB; AAS91342.

XX PT

XX PT New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -

XX PS

XX PS Claim 20; SEQ ID No 57514; 103pp; English.

XX CC

XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG0377 represent novel human diagnostic amino acid sequences of the invention.

XX CC

XX CC Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX SQ

XX SQ Sequence 478 AA;

Query Match

Best Local Similarity 22.1%; Pred. No. 40;

Matches 45; Conservative 20; Mismatches 64; Indels 75; Gaps 10;

Qy 8 ISQSETHAMCWSSSTRSPSRHHLRE-----RIP--CLALGVTAICSVVTH--- 54

Db 78 ITQGSFVTLRCQGSLETO---EYHLVREKKTALWITRPOELVKKGQFFILSITWEHAGR 134

Qy 55 -----VSHGGGISVALCSO-----C 69

Db 135 YCCIYGSHTAGLSESDPLELVVTGAYSKPTLSALPSPVVTSCONTIQCDSQVADGFI 194

Qy 70 LQTNALRPRPDCLTNNGCGYGECHG--SLGHVDRFPQHSNEMN---SGMDSCKP----LR 120

Db 195 LCKEGEDEHPQCLNSHSHARGSSRAIFSVGPV-----SPSRNRSYRCYGYDSRAPYVMSLP 250

Qy 121 GEFLGVLTPI-PRMEPAATRAKGV 143

Db 251 SDLLGLLVPOVKKPSLSVQPGPV 274

RESULT 8

AAU30551

ID AAU30551 standard; Protein; 139 AA.

XX AC

XX AC AAU30551;

XX DT

XX DT 18-DEC-2001 (first entry)

XX DE

XX DE Novel human secreted protein #1042.

XX KW

XX KW Human; vaccination; gene therapy; nutritional supplement; stem cell proliferation; haematopoiesis; nerve tissue regeneration; immune suppression; immune stimulation; anti-inflammatory; leukaemia.

XX OS

XX OS Homo sapiens.

XX PN

XX PN WO200179449-A2.

XX PD

XX PD 25-OCT-2001.

XX PF

XX PF 16-APR-2001; 2001WO-US08656.

XX PR

XX PR 18-APR-2000; 2000US-0552929.

XX PR

XX PR 26-JAN-2001; 2001US-0770160.

XX PA

XX PA (HYSE-) HYSEQ INC.

XX PI

XX PI Tang YT, Liu C, Drmanac RT;

XX DR

XX DR WPI; 2001-611725/70.

XX PT

XX PT Nucleic acids encoding a range of human polypeptides, useful in genetic vaccination, testing and therapy -

XX PS

XX PS Claim 20; Page 309; 765pp; English.

XX CC

XX CC The invention relates to novel human secreted polypeptides. The polypeptides and antibodies to the polypeptides are useful for determining the presence of or predisposition to a disease associated with altered levels of polypeptide. The polypeptides are also useful for identifying agents (agonists and antagonists) that bind to them. Cells expressing the proteins are useful for identifying a therapeutic agent for use in treatment of a pathology related to aberrant expression or physiological interactions of the polypeptide. Vectors comprising the nucleic acids encoding the polypeptides and cells genetically engineered to express them are also useful for producing the proteins. The proteins are useful in genetic vaccination, testing and therapy, and can be used as nutritional supplements. They may be used to increase stem cell proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon and/or nerve tissue growth or regeneration; immune suppression and/or stimulation; as anti-inflammatory agents; and in treatment of leukaemias. AAU29510-AAU3304 represent the amino acid sequences of novel human secreted proteins of the invention.

XX SQ

XX SQ Sequence 139 AA;

Query Match 9.1%; Score 73; DB 22; Length 139;
Best Local Similarity 25.6%; Pred. No. 10;
Matches 30; Conservative 15; Mismatches 42; Indels 30; Gaps 6;

QY 9 SEQETHAWKSS-----STTRSP-----SRHLHREIPCLAGVTAICSL 50
DB 20 SVPARTGGPCWSSSCSNXCTSCSTTPTTLDPGRSHC-----QLPWLKQLPKGMCTS 75

QY 51 VNIHVSHGGGISVALCSQCLQTNALRPRDCLTNNGGCGECHGSLGHVDRFPQHSN 107
DB 76 TCLHGSCG-----CICGSCGPCGI-DCPLGRNHNPHHSHGDLFPY--RFEFAH 124

RESULT 9
AAR29908 standard; Protein; 320 AA.
XX AAR29908;
XX AC AAR29908;
XX DT 25-MAR-2003 (updated)
XX DT 26-APR-1993 (first entry)
XX DE HCV antigen peptide O15-2.
XX KW Clone; Hepatitis C virus; HCV; core-envelope; NS1(gp70); NS2-NS4;
XX KW NS4-NS5; region; diagnostic method; antibody; suppress; control;
XX KW proteolytic; process; precursor; polypeptide.
XX OS Hepatitis C virus.
XX PN EP18313-A2.
XX PD 16-DEC-1992.
XX PF 11-JUN-1992; 92EP-0109812.
XX PR 11-JUN-1991; 91JP-0139268.
XX PR 12-JUL-1991; 91JP-0172794.
XX PR 07-OCT-1991; 91JP-0287008.
XX PR 16-DEC-1991; 91JP-0332329.
XX PR 20-APR-1992; 92JP-0099957.
XX PA (MITU) MITSUBISHI KASEI CORP.
XX PI Hayashi N, Honda Y, Murakami T, Seki M, Takahashi K;
XX PI Teranishi Y;
XX DR WPI; 1992-417213/51.
XX DR N-PSDB; AAQ32539.
XX PT New hepatitis C virus gene and its encoded protein - used for
XX PT diagnosing and vaccinating against hepatitis C virus infections
XX PS Disclosure; Page 277-278; 305pp; English.
XX CC The sequences given in AAR29528 and AAR29907-08 are encoded by various
XX CC clones of the full length Hepatitis C Virus (HCV) gene of the
XX CC invention. These HCV peptides may be useful in the development of a
XX CC diagnostic method which is more accurate and effective than
XX CC conventional ones, in the detection of antibodies raised against a
XX CC wide range of HCVs which have been hardly detected before. The
XX CC complete gene may be used in an in vitro screening system for a
XX CC substance capable of specifically suppressing or controlling a
XX CC proteolytic processing of a precursor polypeptide of HCV. See also
XX CC AAQ32436.
XX CC (Updated on 25-MAR-2003 to correct PN field.)
XX SQ Sequence 320 AA;

Query Match 9.1%; Score 73; DB 13; Length 320;
Best Local Similarity 23.0%; Pred. No. 28;
Matches 28; Conservative 21; Mismatches 45; Indels 28; Gaps 5;

QY 30 HLHREIPCLAGVTAICSLVNIHVSHGGGISVALCSQCLQTN-----LRPRDC 81
DB 40 HPIITKFMACNSADLEVVTS-TWLV---GGVLAALAAAYCLTTGSVIVIGRIILSGRP 95

QY 82 LTNNGGCV-----GECHGSLGHVDRFPQHSNENKNSGMDCKPLRGSEFLGVLTPHPKMEF 135
DB 96 IPDREVLRYREFDEMEECASHLPYIEQ-----GNQLAEQPKQKALGLTTATKQAE 145

QY 136 AA 137
DB 146 AA 147

RESULT 10
AAW99482 standard; Protein; 1463 AA.
XX AAW99482;
XX AC AAW99482;
XX DT 08-JUN-1999 (first entry)
XX DE Murine NCoA-2 protein.
XX KW Mouse; p/CIP; p300/CBP/co-integrator-associated protein; gene expression;
XX KW breast cancer; inflammatory disease; atherosclerosis; osteoporosis.
XX OS Mus sp.
XX PH Key Location/Qualifiers
XX FT Misc-difference 989 /label= unknown
XX PN WO9856806-A1.
XX PD 17-DEC-1998.
XX PR 12-JUN-1998; 98WO-US12263.
XX PR 12-JUN-1997; 97US-0049452.
XX PA (REGC) UNIV CALIFORNIA.
XX PI Glass CK, Rose DW, Rosenfield MG, Torchia J;
XX PI WPI; 1999-080883/07.
XX PT New nucleic acids encoding p/CIP and NCoA-2 polypeptides - are used
XX PT to identify agents that regulate gene expression, e.g. for treatment
XX PT of cancer, inflammatory disease and osteoporosis
XX PS Claim 22; Fig 2A; 100pp; English.
XX CC This sequence represents the amino acid sequence of the mouse NCoA-2.
XX CC The protein can regulate gene expression so are potentially useful
XX CC therapeutically, e.g. against (breast) cancer, inflammatory disease
XX CC (e.g. atherosclerosis) or osteoporosis.
XX SQ Sequence 1463 AA;

Query Match 9.1%; Score 73; DB 20; Length 1463;
Best Local Similarity 25.3%; Pred. No. 1.7e+02;
Matches 39; Conservative 12; Mismatches 43; Indels 60; Gaps 8;

QY 21 SSTTRSP-----SRHLHREIPCL-----ALGVTAICS----- 49
DB 358 SQTTFEPQLVLSLHLHREQVCMVNPDLTGQANGKPLNFISSSSPAHQALCSGNPQGDH 417

QY 50 --LVNIHVSHGGGISVALCSQCLQTNALRPRDCLTNNG--GCYGECHGSLGHVDRFPQH 105
DB 418 TLSSHHIFPHNG-----PKQHQHGHPRFGSGGMVSGVGHQATTPQG 460

QY 106 SN---EWNKSGMDSK---KPLRGEFLGVLTPHPK 133

Db 461 SNYALKONSPQSSPGHNP--GQASSVLSPQRQ 492

RESULT 11

AAV19454

ID AAV19454 standard; Protein; 131 AA.

XX AAV19454;

AC AAV19454;

XX 14-JUL-1999 (first entry)

DT 14-JUL-1999 (first entry)

DE Amino acid sequence of a human secreted protein.

XX Human secreted protein; cancer; tumour; neurodegenerative disorder;

KW developmental abnormality; fetal deficiency; blood disorder; leukemia;

KW immune system disease; autoimmune disease; hepatic disease;

KW renal disease; lymphoma; inflammation; allergy; ischemic shock;

KW Alzheimer's; cognitive disorder; schizophrenia; prostate disease;

KW obesity; osteoclast; osteoporosis; arthritis; malignancy; testes disease;

KW lung disease; thymus disease; digestive disorder; endocrine disorder;

KW infection; AIDS.

XX Homo sapiens.

OS Homo sapiens.

XX WO922243-A1.

PN 06-MAY-1999.

XX 23-OCT-1998; 98MO-US22376.

XX 24-OCT-1997; 97US-0063387.

PR 24-OCT-1997; 97US-0063784.

PR 24-OCT-1997; 97US-0063088.

PR 24-OCT-1997; 97US-0063089.

PR 24-OCT-1997; 97US-0063090.

PR 24-OCT-1997; 97US-0063091.

PR 24-OCT-1997; 97US-0063092.

PR 24-OCT-1997; 97US-0063097.

PR 24-OCT-1997; 97US-0063098.

PR 24-OCT-1997; 97US-0063099.

PR 24-OCT-1997; 97US-0063100.

PR 24-OCT-1997; 97US-0063101.

PR 24-OCT-1997; 97US-0063109.

PR 24-OCT-1997; 97US-0063110.

PR 24-OCT-1997; 97US-0063111.

PR 24-OCT-1997; 97US-0063148.

PR 24-OCT-1997; 97US-0063386.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Brewer LA, Carter KC, Duan DR, Ebner R, Endress GA;

PI Feng P, Florence C, Florence KA, Greene JM, Janat F;

PI Kay H, Lafleur DM, Moore PA, Ni J, Olsen HS, Rosen CA;

PI Ruben SM, Shi Y, Soppet DR, Wei Y, Young P;

XX WPI; 1999-303069/25.

DR N-PSDB; AAX61334.

XX New isolated human genes and the secreted polypeptides they encode

PT Claim 11; Page 393; 546pp; English.

XX The specification describes cDNA sequences (AAX61322-X61470) encoding

CC human secreted proteins (AAV19442-Y19590). The polynucleotides and their

CC corresponding secreted polypeptides are useful for preventing, treating

CC or ameliorating medical conditions, e.g. by protein or gene therapy.

CC Pathological conditions can also be diagnosed by determining the amount

CC of the polypeptides in a sample or by determining the presence of

CC mutations in the polynucleotides. Specific uses are described for each

CC of the polynucleotides, based on which tissues they are most highly

CC expressed in, and include developing products for the diagnosis or

CC treatment of cancer, tumours, neurodegenerative disorders, developmental

CC abnormalities and fetal deficiencies, blood disorders, leukemias,

CC diseases of the immune system, autoimmune diseases, hepatic and renal

CC disease, lymphomas, inflammation, allergies, ischemic shock, Alzheimer's

CC and cognitive disorders, schizophrenia, prostate diseases, obesity,

CC disorders involving osteoclasts such as osteoporosis, arthritis or

CC malignancies, diseases of testes, lung or thymus, digestive/endocrine

CC disorders, infections and AIDS. The polypeptides are also useful for

CC identifying their binding partners.

XX Sequence 131 AA;

SQ Query Match 8.8; Score 71; DB 20; Length 131;

Best Local Similarity 33.3; Pred. NO. 16;

Matches 26; Conservative 4; Mismatches 40; Indels 8; Gaps 3;

QY 19 WSSSTTRSPSR-HLHRRERIPCLALGVTAI--CSLYMIVH-----SHGGGIVSVALCSQCL 70

DB 2 WSVIRSLCPRLQSLHVCFCPRCLAVPCVPHLLSSPWFHVRXXAFSXPCCIGICFVGL 61

QY 71 QTNALRRPDLCTNNGGC 88

DB 62 LLGAXRPRSGCLCSPSXC 79

RESULT 12

AAU31738

ID AAU31738 standard; Protein; 286 AA.

XX AAU31738;

XX 18-DEC-2001 (first entry)

DT Novel human secreted protein #2229.

DE Human; vaccination; gene therapy; nutritional supplement;

XX stem cell proliferation; haematopoiesis; nerve tissue regeneration;

KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.

XX Homo sapiens.

XX WO200179449-A2.

PN 25-OCT-2001.

XX 16-APR-2001; 2001WO-US08656.

XX 18-APR-2000; 2000US-0552929.

PR 26-JAN-2001; 2001US-0770160.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-611725/70.

XX Nucleic acids encoding a range of human polypeptides, useful in genetic

PT vaccination, testing and therapy -

XX Claim 20; Page 502; 765pp; English.

XX The invention relates to novel human secreted polypeptides. The

CC polypeptides and antibodies to the polypeptides are useful for

CC determining the presence of or predisposition to a disease associated

CC with altered levels of polypeptide. The polypeptides are also useful for

CC identifying agents (agonists and antagonists) that bind to them. Cells

CC expressing the proteins are useful for identifying a therapeutic agent

CC for use in treatment of a pathology related to aberrant expression or

CC physiological interactions of the polypeptide. Vectors comprising

CC the nucleic acids encoding the polypeptides and cells genetically

CC engineered to express them are also useful for producing the proteins.

CC The proteins are useful in genetic vaccination, testing and

CC therapy, and can be used as nutritional supplements. They may be used to

CC increase stem cell proliferation; to regulate haematopoiesis; and in

CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;

CC immune suppression and/or stimulation; as anti-inflammatory agents; and
CC in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid
CC sequences of novel human secreted proteins of the invention.

XX
SQ Sequence 286 AA;

Query Match 8.8%; Score 71; DB 22; Length 286;
Best Local Similarity 22.6%; Pred. No. 39;
Matches 28; Conservative 10; Mismatches 62; Indels 24; Gaps 4;

Oy 23 TTRSPSRHLLHREIPCLALGVTAICLSLVWTHV-----SHGGGISVALCSQCLQTNALRRP 78
Db 73 TNSPPPHHTLPQRPYPPTGX-----HTQASHHPAGKXASSQPRASPASRPS 124

Oy 79 PDCLTNGGCGECHGS-----LGHVDRFPQHNEWNSGMDSC-----KPLRGEFLGV 126
Db 125 PTDALRANPLRSVSGSLPTFPYLHCNMPKACSPWRPAADMTARPRPRPPKDPQGP 184

Oy 127 LTPH 130
Db 185 ARAH 188

RESULT 13
AAE00913
ID AAE00913 standard; Protein; 1686 AA.
AC AC
XX AAE00913;
DT 04-JUL-2001 (first entry)
DE Human 27875 ADAM-TS protein, alternative version.
XX Human; ADAM-TS; A Disintegrin And Metalloproteinase; antiinflammatory;
KW angio genesis; bronchial asthma; Goodpasture's syndrome; metastasis;
KW heart failure; cardiac hypertrophy; chronic ischaemic heart disease;
KW sickle cell disease nephropathy; urinary tract obstruction; haemostatic;
KW skeletal muscle disorder; myocardial infarction; blood vessel disorder;
KW hypertension; atherosclerosis; vasculitis; renal artery stenosis; mumps;
KW gonorrhea; tuberculosis; syphilis; spermatocytic seminoma; osteoporosis;
KW rhabdomyosarcoma; glomerulonephritis; bone disorder; Paget's disease;
KW rickets; osteomalacia; Hodgkin's disease; gene therapy; antibacterial;
KW candidant; tumour; thymoma; vasotropic; cytostatic; virucide.
XX Homo sapiens.
OS
XX
XX
PH Key
FT Peptide
FT 1..30
FT /label= Signal_peptide
FT 6..8
FT /notes= "Protein kinase C phosphorylation site"
FT Protein
FT 31..1686
FT /label= Human_mature_27875_ADAM-TS_protein
FT 55..60
FT /notes= "N-myristoylation site"
FT Modified-site
FT 73..75
FT /notes= "protein kinase C phosphorylation site"
FT Domain
FT 78..93
FT /notes= "Crystallins beta and gamma Greek key motif"
FT Modified-site
FT 94..97
FT /notes= "N-glycosylation site"
FT Modified-site
FT 110..112
FT /notes= "Protein kinase C phosphorylation site"
FT Modified-site
FT 115..120
FT /notes= "N-myristoylation site"
FT Modified-site
FT 141..146
FT /notes= "N-myristoylation site"
FT Modified-site
FT 147..150
FT /notes= "Casein kinase II phosphorylation site"
FT Modified-site
FT 159..162
FT /notes= "Casein kinase II phosphorylation site"
FT Binding-site
FT 195..197
FT /note= "Cell attachment sequence"

FT Modified-site
FT 214..217
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site
FT 214..216
FT /note= "Protein kinase C phosphorylation site"
FT Domain
FT 244..259
FT /label= Metalloproteinase_domain
FT Modified-site
FT 313..315
FT /note= "Protein kinase C phosphorylation site"
FT Modified-site
FT 342..345
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site
FT 342..344
FT /note= "Protein kinase C phosphorylation site"
FT Modified-site
FT 373..376
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site
FT 379..384
FT /note= "N-myristoylation site"
FT Domain
FT 385..394
FT /label= Zinc_binding_domain
FT Modified-site
FT 401..404
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site
FT 408..411
FT /note= "Amidation site"
FT Modified-site
FT 479..484
FT /note= "N-myristoylation site"
FT Domain
FT 488..567
FT /label= Thrombospondin_domain
FT Modified-site
FT 505..508
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site
FT 513..518
FT /note= "N-myristoylation site"
FT Domain
FT 539..545
FT /note= "Growth factor and cytokine metalloproteinase family signature 2 domain"
FT Modified-site
FT 539..544
FT /note= "N-myristoylation site"
FT Domain
FT 541..592
FT /label= Disintegrin_domain
FT Domain
FT 542..592
FT /label= Thrombospondin_domain
FT Modified-site
FT 557..562
FT /note= "N-myristoylation site"
FT Modified-site
FT 569..571
FT /note= "Protein kinase C phosphorylation site"
FT Misc-difference
FT 596
FT /label= Unknown
FT /note= "Encoded by CRC"
FT Modified-site
FT 598..600
FT /note= "Protein kinase C phosphorylation site"
FT Modified-site
FT 605..608
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site
FT 614..619
FT /note= "N-myristoylation site"
FT Modified-site
FT 667..672
FT /note= "N-myristoylation site"
FT Binding-site
FT 687..692
FT /note= "Cytochrome C family heme-binding site"
FT Modified-site
FT 688..693
FT /note= "N-myristoylation site"
FT Modified-site
FT 693..696
FT /note= "N-glycosylation site"
FT Modified-site
FT 703..706
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site
FT 716..721
FT /note= "N-myristoylation site"
FT Modified-site
FT 740..747
FT /note= "Tyrosine kinase phosphorylation site"
FT Modified-site
FT 765..770
FT /note= "N-myristoylation site"
FT Modified-site
FT 774..779
FT /note= "N-myristoylation site"
FT Modified-site
FT 778..781
FT /note= "N-glycosylation site"
FT Domain
FT 825..868


```

PT /label= Thrombospondin_domain
PT 872..875
PT Modified-site
PT /note= "cAMP- and cGMP-dependent protein kinase
PT phosphorylation site"
PT 901..903
PT Modified-site
PT /note= "Protein kinase C phosphorylation site"
PT 917..920
PT Modified-site
PT /note= "Casein kinase II phosphorylation site"
PT 949..994
PT Domain
PT /label= Thrombospondin_domain
PT 949..988
PT Domain
PT /label= Thrombospondin_domain
PT 950..953
PT Modified-site
PT /note= "N-glycosylation site"
PT 957..960
PT Modified-site
PT /note= "Casein kinase II phosphorylation site"
PT 962..964
PT Modified-site
PT /note= "Protein kinase C phosphorylation site"
PT 971..974
PT Modified-site
PT /note= "N-glycosylation site"
PT 1005..1010
PT Modified-site
PT /note= "N-myristoylation site"
PT 1006..1009
PT Binding-site
PT /note= "Glycosaminoglycan attachment site"
PT 1011..1014
PT Modified-site
PT /note= "Casein kinase II phosphorylation site"
PT 1035..1037
PT Modified-site
PT /note= "Protein kinase C phosphorylation site"
PT 1039..1044
PT Modified-site
PT /note= "N-myristoylation site"
PT 1192..1195
PT Modified-site
PT /note= "Casein kinase II phosphorylation site"
PT 1263..1252
PT Modified-site
PT /note= "N-myristoylation site"
PT 1263..1268
PT Modified-site
PT /note= "N-myristoylation site"
PT 1308..1311
PT Modified-site
PT /note= "Casein kinase II phosphorylation site"
PT 1358..1363
PT Modified-site
PT /note= "N-myristoylation site"
PT 1370..1372
PT Modified-site
PT /note= "Protein kinase C phosphorylation site"
PT 1385..1387
PT Modified-site
PT /note= "Protein kinase C phosphorylation site"
PT 1397..1400
PT Modified-site
PT /note= "Casein kinase II phosphorylation site"
PT 1412..1415
PT Modified-site
PT /note= "N-glycosylation site"
PT 1415..1463
PT Domain
PT /label= Thrombospondin_domain
PT 1419..1422
PT Modified-site
PT /note= "N-glycosylation site"
PT 1440..1443
PT Modified-site
PT /note= "Casein kinase II phosphorylation site"
PT 1440..1442
PT Modified-site
PT /note= "Protein kinase C phosphorylation site"
PT 1466..1521
PT Domain
PT /label= Thrombospondin_domain
PT 1470..1473
PT Modified-site
PT /note= "N-glycosylation site"
PT 1483..1486
PT Modified-site
PT /note= "Casein kinase II phosphorylation site"
PT 1483..1485
PT Modified-site
PT /note= "Protein kinase C phosphorylation site"
PT 1517..1522
PT Modified-site
PT /note= "N-myristoylation site"
PT 1528..1531
PT Modified-site
PT /note= "Casein kinase II phosphorylation site"
PT 1528..1530
PT Modified-site
PT /note= "Protein kinase C phosphorylation site"
PT 1546..1549
PT Modified-site
PT /note= "Casein kinase II phosphorylation site"
PT 1592..1597

```

```

PT /note= "N-myristoylation site"
PT 1599..1601
PT Modified-site
PT /note= "Protein kinase C phosphorylation site"
PT 1606..1609
PT Modified-site
PT /note= "cAMP- and cGMP-dependent protein kinase
PT phosphorylation site"

Query Match      8.8; Score 71; DB 22; Length 1686;
Best Local Similarity 23.1%; Pred. No. 3.2e+02; Indels 50; Gaps 8;
Matches 31; Conservative 14; Mismatches 39;

QV 27 PSRHLHRIPLALGVTAICSLVMIHVSH--GGG-----ISVALCSQCLQTNALR 76
Db 1507 PAKPHAHR---PCGAQPCLSWYTSWRCEACGGGQQRLVTCPEPGICEALPNTTR 1563
QV 77 PRPDCLTNN-----GGCYGEGHGSLS-----GHVDRFPCH 105
Db 1564 P---CNTHPTQWVVGPGQCSAPCGGVORRLVKVNTQTGLPREDSQCGH-EAWPES 1619
QV 106 SNEWMNGMDSCKPL 119
Db 1620 SRP--CGTEDCPEV 1631

RESULT 14
AAE00934
ID AAE00934 standard; Protein; 1686 AA.
XX AC AAE00934;
XX DT 04-JUL-2001 (first entry)
XX DE Human 27875 ADAM-TS (a disintegrin and metalloproteinase).
XX KW Human; ADAM-TS; A Disintegrin And Metalloproteinase; antiinflammatory;
XX KW angiogenesis; bronchial asthma; Goodpasture's syndrome; metastasis;
XX KW heart failure; cardiac hypertrophy; chronic ischaemic heart disease;
XX KW sickle cell disease nephropathy; urinary tract obstruction; haemostatic;
XX KW skeletal muscle disorder; myocardial infarction; blood vessel disorder;
XX KW hypertension; atherosclerosis; vasculitis; renal artery stenosis; mumps;
XX KW gonorrhea; tuberculosis; syphilis; spermatocytic seminoma; osteoporosis;
XX KW rhabdomyosarcoma; glomerulonephritis; bone disorder; Paget's disease;
XX KW rickets; osteomalacia; Hodgkin's disease; gene therapy; antibacterial;
XX KW cardiac; tumour; thymoma; vasotropic; cytostatic; virucide.
XX CS Homo sapiens.
XX FH
XX Key Location/Qualifiers
PT Peptide 1..30
PT /label= Signal_peptide
PT Modified-site 6..8
PT /note= "Protein kinase C phosphorylation site"
PT Protein 31..1686
PT /label= Human_mature_27875_ADAM-TS_protein
PT Modified-site 55..60
PT /note= "N-myristoylation site"
PT Modified-site 73..75
PT /note= "Protein kinase C phosphorylation site"
PT Domain 78..93
PT /note= "Crystallins beta and gamma Greek key motif"
PT Modified-site 94..97
PT /note= "N-glycosylation site"
PT Modified-site 110..112
PT /note= "Protein kinase C phosphorylation site"
PT Modified-site 115..120
PT /note= "N-myristoylation site"
PT Modified-site 141..146
PT /note= "N-myristoylation site"
PT Modified-site 147..150
PT /note= "Casein kinase II phosphorylation site"
PT Modified-site 159..162
PT /note= "Casein kinase II phosphorylation site"
PT Binding-site 195..197

```

PT	Modified-site	/note= "Cell attachment sequence"	FT	Modified-site	/note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"
PT	Modified-site	214..217	FT	Modified-site	901..903
PT	Modified-site	/note= "Casein kinase II phosphorylation site"	PT	Modified-site	/note= "Protein kinase C phosphorylation site"
PT	Domain	214..216	PT	Domain	917..920
PT	Modified-site	/note= "Protein kinase C phosphorylation site"	PT	Domain	/note= "Casein kinase II phosphorylation site"
PT	Modified-site	244..259	PT	Domain	/label= Thrombospondin_domain
PT	Modified-site	/label= Metalloproteinase_domain	PT	Domain	/label= Thrombospondin_domain
PT	Modified-site	313..315	PT	Modified-site	/note= "N-glycosylation site"
PT	Modified-site	/note= "Protein kinase C phosphorylation site"	PT	Modified-site	950..953
PT	Modified-site	342..345	PT	Modified-site	957..960
PT	Modified-site	/note= "Casein kinase II phosphorylation site"	PT	Modified-site	/note= "Casein kinase II phosphorylation site"
PT	Modified-site	342..344	PT	Modified-site	962..964
PT	Modified-site	/note= "Protein kinase C phosphorylation site"	PT	Modified-site	/note= "Protein kinase C phosphorylation site"
PT	Modified-site	373..376	PT	Modified-site	971..974
PT	Modified-site	/note= "Casein kinase II phosphorylation site"	PT	Modified-site	/note= "N-glycosylation site"
PT	Modified-site	379..384	PT	Modified-site	1005..1010
PT	Domain	/note= "N-myristoylation site"	PT	Modified-site	/note= "N-myristoylation site"
PT	Modified-site	385..394	PT	Modified-site	1006..1009
PT	Modified-site	/label= Zinc_binding_domain	PT	Modified-site	/note= "Glycosaminoglycan attachment site"
PT	Modified-site	401..404	PT	Modified-site	1011..1014
PT	Modified-site	/note= "Casein kinase II phosphorylation site"	PT	Modified-site	/note= "Casein kinase II phosphorylation site"
PT	Modified-site	408..411	PT	Modified-site	1035..1037
PT	Modified-site	/note= "Amidation site"	PT	Modified-site	/note= "Protein kinase C phosphorylation site"
PT	Modified-site	479..484	PT	Modified-site	/note= "N-myristoylation site"
PT	Domain	/note= "N-myristoylation site"	PT	Modified-site	1039..1044
PT	Modified-site	488..567	PT	Modified-site	1192..1195
PT	Modified-site	/label= Thrombospondin_domain	PT	Modified-site	/note= "Casein kinase II phosphorylation site"
PT	Modified-site	505..508	PT	Modified-site	1263..1268
PT	Modified-site	/note= "Casein kinase II phosphorylation site"	PT	Modified-site	/note= "N-myristoylation site"
PT	Modified-site	513..518	PT	Modified-site	1308..1311
PT	Domain	/note= "N-myristoylation site"	PT	Modified-site	/note= "Casein kinase II phosphorylation site"
PT	Modified-site	539..545	PT	Modified-site	1358..1363
PT	Modified-site	/note= "Growth factor and cytokine metalloproteinase family signature 2 domain"	PT	Modified-site	/note= "N-myristoylation site"
PT	Domain	/note= "N-myristoylation site"	PT	Modified-site	1370..1372
PT	Modified-site	541..592	PT	Modified-site	/note= "Protein kinase C phosphorylation site"
PT	Domain	/label= Disintegrin_domain	PT	Modified-site	1385..1387
PT	Modified-site	542..592	PT	Modified-site	/note= "Protein kinase C phosphorylation site"
PT	Modified-site	/label= Thrombospondin_domain	PT	Modified-site	1397..1400
PT	Modified-site	557..562	PT	Modified-site	/note= "Casein kinase II phosphorylation site"
PT	Modified-site	/note= "N-myristoylation site"	PT	Modified-site	1412..1415
PT	Modified-site	569..571	PT	Domain	/note= "N-glycosylation site"
PT	Modified-site	/note= "Protein kinase C phosphorylation site"	PT	Modified-site	1415..1463
PT	Modified-site	598..600	PT	Modified-site	/label= Thrombospondin_domain
PT	Modified-site	/note= "Protein kinase C phosphorylation site"	PT	Modified-site	1419..1422
PT	Modified-site	605..608	PT	Modified-site	/note= "N-glycosylation site"
PT	Modified-site	/note= "Casein kinase II phosphorylation site"	PT	Modified-site	1440..1443
PT	Modified-site	614..619	PT	Modified-site	/note= "Casein kinase II phosphorylation site"
PT	Modified-site	/note= "N-myristoylation site"	PT	Domain	1466..1521
PT	Modified-site	667..672	PT	Modified-site	/label= Thrombospondin_domain
PT	Modified-site	/note= "N-myristoylation site"	PT	Modified-site	1470..1473
PT	Modified-site	687..692	PT	Modified-site	/note= "N-glycosylation site"
PT	Modified-site	/note= "Cytochrome C family heme-binding site"	PT	Modified-site	1483..1486
PT	Modified-site	688..693	PT	Modified-site	/note= "Casein kinase II phosphorylation site"
PT	Modified-site	/note= "N-myristoylation site"	PT	Modified-site	1483..1485
PT	Modified-site	693..696	PT	Modified-site	/note= "Protein kinase C phosphorylation site"
PT	Modified-site	/note= "N-glycosylation site"	PT	Modified-site	1517..1522
PT	Modified-site	703..706	PT	Modified-site	/note= "N-myristoylation site"
PT	Modified-site	/note= "Casein kinase II phosphorylation site"	PT	Modified-site	1528..1531
PT	Modified-site	716..721	PT	Modified-site	/note= "Casein kinase II phosphorylation site"
PT	Modified-site	/note= "N-myristoylation site"	PT	Modified-site	1528..1530
PT	Modified-site	740..747	PT	Modified-site	/note= "Protein kinase C phosphorylation site"
PT	Modified-site	/note= "Tyrosine kinase phosphorylation site"	PT	Modified-site	1546..1549
PT	Modified-site	765..770	PT	Modified-site	/note= "Casein kinase II phosphorylation site"
PT	Modified-site	/note= "N-myristoylation site"	PT	Modified-site	1592..1597
PT	Modified-site	774..779	PT	Modified-site	/note= "N-myristoylation site"
PT	Modified-site	/note= "N-myristoylation site"	PT	Modified-site	1599..1601
PT	Modified-site	778..781			
PT	Domain	/note= "N-glycosylation site"			
PT	Modified-site	825..868			
PT	Modified-site	/label= Thrombospondin_domain			
PT	Modified-site	872..875			

FT Modified-site /note= "protein kinase C phosphorylation site"
 1506...1609
 FT /note= "cAMP- and cGMP-dependent protein kinase
 phosphorylation site"
 1620...1622
 FT Modified-site /note= "protein kinase C phosphorylation site"
 1625...1630
 FT Modified-site
 Query Match 8.8%; Score 71; DB 22; Length 1586;
 Best Local Similarity 23.1%; Pred. No. 3.2e+02;
 Matches 31; Conservative 14; Mismatches 39; Indels 50; Gaps 8;
 QY 27 PSRHLHRRIPCLALGVTAICSLVHSH--GGG-----ISVALCSQCLQTNALR 76
 DB 1507 PAKPAHR---PCGAQPCLSWTSSWRECEACGGGQQRLVTCPEGLCEBALRNTR 1563
 QY 77 PRPDCLTNN-----GGCGEGHGSLSL-----CHVDRFPQH 105
 DB 1564 P---CNTHCTQWVGPWCGGAPCGGVORLVKCVNTQTGLPBEDSQOQH-SAMPES 1619
 QY 106 SNEWNSGMDSCKPL 119
 DB 1620 SRP--CGTEDCEPV 1631

RESULT 15
 ABG26623
 ID ABG26623 standard; Protein: 143 AA.

AC ABG26623;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #26614.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WC-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.

DR N-PSDB; AAS90310.

XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -

PS Claim 20; SEQ ID No 56982; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from Wipo
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 143 AA;

Query Match 8.8%; Score 70.5; DB 22; Length 143;

Best Local Similarity 25.0%; Pred. No. 20;

Matches 36; Conservative 12; Mismatches 69; Indels 27; Gaps 7;

QY 7 PISQSE---THAWCWSSTTRPSRRHHLHRRIPCLALGVTAICSLVHSHGGGSI 62

DB 6 PLSABSRPEGVRLFC---SOFASRQPPQAKS-----AWGTTALSESMMCFASFGASIS 57

QY 63 VALCSQCLQTNALRPRPDCLTNNGCCYGCCHSLGHVDR-----FPOHSNENNSQMDSCKP 118

DB 58 SSADQEQHL-----PGAV--GSGRPGECIGPSGRCSKXGATFPGOKGRWKEARPSVP 109

QY 119 LRGEFLGLVTPHPKPEFAAIRAGK 142

DB 110 APAAGLOSRAHPR---GGIRPCR 130

Search completed: November 13, 2003, 13:19:39

Job time : 42 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 14, 2003, 12:21:59 ; Search time 1916 Seconds
(without alignments)
1813.956 Million cell updates/sec

Title: US-09-868-025-2

Perfect score: 803

Sequence: 1 MVDNFPISQSETHACWS.....LGVLPHPKMEFAAIRAGKV 143

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n_model_DEV.xlh
-Q=/cgn2_1/USPFO_spool/US09868025/runat_13112003_130406_17467/app_query.fasta_1.327
-DB=EST -QFMT=fastap -SUFFIX=first -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCLIGN=200 -THR SCORE=500 -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09868025 -CCN_1_1_2135 -runat_13112003_130406_17467 -NCPU=6 -ICPU=3
-NO_WMAP -LARGESQUERY -NEG_SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:
1: em_eatba.*
2: em_esthum.*
3: em_estcin.*
4: em_estm.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_hic.*
9: gb_est1.*
10: gb_est2.*
11: gb_hic.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_pln.*
20: em_gss_vrt.*
21: em_gss_vrt.*
22: em_gss_fun.*
23: em_gss_mam.*
24: em_gss_mue.*
25: em_gss_pro.*
26: em_gss_rpd.*
27: em_gss_phg.*
28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	744	92.7	723	9	AU068433	AU068433 OSJNEC08C
C 2	744	92.7	727	14	CB655085	CB655085 OSJNEC08C
C 3	744	92.7	763	14	CB663774	CB663774 OSJNEC08C
C 4	744	92.7	776	14	CB000367	CB000367 S345U_H05
C 5	744	92.7	777	14	CB627453	CB627453 OSJNEC08C
C 6	744	92.7	780	14	CB655151	CB655151 OSJNEC08C
C 7	744	92.7	792	14	CB647160	CB647160 OSJNEC08C
C 8	744	92.7	799	14	CB620522	CB620522 OSJNEC08C
C 9	744	92.7	805	14	CB660999	CB660999 OSJNEC08C
C 10	744	92.7	816	14	CB633818	CB633818 OSJNEC08C
C 11	744	92.7	836	14	CB659692	CB659692 OSJNEC08C
C 12	744	92.7	850	14	CB619892	CB619892 OSJNEC08C
C 13	740	92.2	682	14	CB647623	CB647623 OSJNEC08C
C 14	737	91.8	653	14	CB643278	CB643278 OSJNEC08C
C 15	731	91.0	712	9	AU075373	AU075373 AU075373
C 16	727	90.5	705	9	AU165747	AU165747 AU165747
C 17	717	89.3	768	14	CB649882	CB649882 OSJNEC08C
C 18	684	85.2	718	9	AU031690	AU031690 AU031690
C 19	682	84.9	622	13	BQ06858	BQ06858 MD20F01 O
C 20	631.5	78.6	686	9	AU165429	AU165429 AU165429
C 21	601	74.8	419	14	D48832	D48832 RICS15269A
C 22	558	69.5	456	9	AT003421	AT003421 AT003421
C 23	481	59.3	321	10	BE230720	BE230720 99MJ117 R
C 24	479	59.7	464	9	AU184049	AU184049 AU184049
C 25	406	50.6	459	9	AT003640	AT003640 AT003640
C 26	399	49.7	691	10	BE229010	BE229010 98BS0003
C 27	394.5	49.1	510	12	BM419698	BM419698 R017808 O
C 28	392	48.8	449	9	AT003707	AT003707 AT003707
C 29	381	47.4	484	12	BI797263	BI797263 H070D10 E
C 30	369.5	46.0	713	12	BM078268	BM078268 HEST117 C
C 31	368.5	45.9	605	13	BQ172639	BQ172639 10910250C
C 32	359.5	44.8	530	14	CA617150	CA617150 w11n.pk00
C 33	357	44.5	675	13	BQ068661	BQ068661 WHE3584 A
C 34	356	44.3	557	12	BJ322180	BJ322180 BJ322180
C 35	355.5	44.3	560	9	AT1737464	AT1737464 606040H06
C 36	355.5	44.3	591	14	CA759397	CA759397 BR060C08B
C 37	354	44.1	587	9	AV927466	AV927466 AV927466
C 38	354	44.1	587	9	AV927464	AV927464 AV927464
C 39	354	44.1	608	9	AV927804	AV927804 AV927804
C 40	354	44.1	651	12	BJ468238	BJ468238 BJ468238
C 41	354	44.1	675	9	AV927460	AV927460 AV927460
C 42	354	44.1	691	12	BJ475281	BJ475281 BJ475281
C 43	354	44.1	695	9	AV931566	AV931566 AV931566
C 44	354	44.1	729	12	BJ468250	BJ468250 BJ468250
C 45	353	44.0	659	12	BJ321932	BJ321932 BJ321932

ALIGNMENTS

RESULT 1
LOCUS AU068433/c
DEFINITION AU068433 Rice callus Oryza sativa (japonica cultivar-group) cDNA
Clone C0250_52, mRNA sequence.
ACCESSION AU068433
VERSION AU068433.1 GI:5003284
KEYWORDS EST.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Zukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

REFERENCE 1 (bases 1 to 723)
 AUTHORS Yamamoto, K. and Sasaki, T.
 TITLE Rice cDNA from callus (1998)
 JOURNAL Unpublished
 COMMENT Contact: Takuji Sasaki
 National Institute of Agrobiological Resources
 Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
 305-8602, Japan
 Tel: 81-298-38-7441
 Fax: 81-298-38-7468
 Email: tsasaaki@affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/
 PROJECT = 'RGP'

FEATURES
 source
 1..723
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="C30250.52"
 /clone_lib="Rice callus"
 /notes="Vector: pBluescript II SK+; Site 1: SalI; Site 2:
 NotI; cDNA prepared from rice callus mRNAs by using
 oligo(dT) as a primer and ligating to the SalI-NotI site
 of pBluescript II SK+ phagemid."
 BASE COUNT 183 a 154 c 183 g 203 t

ORIGIN
 Alignment Scores:
 Pred. No.: 4,15e-63 Length: 723
 Score: 744.00 Matches: 134
 Percent Similarity: 97.84% Conservative: 2
 Best Local Similarity: 96.40% Mismatches: 2
 Query Match: 92.65% Indels: 1
 DB: 9 Gaps: 0

US-09-868-025-2 (1-143) x AU068433 (1-723)

QY 1 MetValAspThrAsnProIleSerGluGlnSerGluThrHisAlaTrpCysTrpSer 20
 Db CTGTTAGATACAAATTTTCCAAATCAGTGAGCAATCAGAAACACATGCTTGGTGTGGAGC 399

QY 21 SerSerThrThrArgSerProSerArgHisLeuHisArgGluArgIleProCysLeu 40
 Db AGCAGACACACAGAGCCGCTCCGCCATCACTTCATCGAGCGGATCCCATGCTT 339

QY 41 AlaLeuGlyValThrAlaIleCysSerLeuValTrpIleHisValSerHisGlyGly 60
 Db GCCCTCGGTGTCACTGCAATTTCCAGTCTTGTGGATACATGTCAGCCACGGTGTGGA 279

QY 61 IleSerValAlaLeuCysSerGlnCysLeuGlnThrAsnAlaLeuArgProAsp 80
 Db ATCAGCGTGCCTTTTCAGCCAGTGCCTTCAGACGATGCGTCCGCCCTCGACCGAT 219

QY 81 CysLeuThrAsnAsnGlyGlyCysTyrglyCysHisGlySerLeuGlyHisValAsp 100
 Db TGTCTGACCAACAATGGAGGCTGCTATGCTGAATGCCATGCCAGTCTTGGCGCATGTGAC 159

QY 101 ArgPheProGlnHisSerAsnGluTrpAsnSerGlyMetAspSerCysValProLeuArg 120
 Db AGATTTCTCAGCATTCATATGAATGAACAGCGGCGATGACAGCTGCCAACCACTGAGG 99

QY 121 GlyGluPheLeuGlyValLeuThrProHisProIleProIleMetGluPheAlaAlaIle 138
 Db GGAGAAATTTCTTGGTGTGCTCAGCGCCACACCCCAAGAGGATGTTTCTGCCATT 44

RESULT 2
 CB655085 727 bp mRNA linear EST 09-APR-2003
 LOCUS OSJNEC08C12.3 rXJNEC Oryza sativa (japonica cultivar-group) cDNA
 DEFINITION clone OSJNEC08C12.3, mRNA sequence.
 ACCESSION CB655085
 VERSION CB655085.1 GI:29659810
 KEYWORDS EST.

SOURCE
 ORGANISM Oryza sativa (japonica cultivar-group)
 Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 REFERENCE 1 (bases 1 to 723)
 AUTHORS Jantassurayarat, C., Lu, G., Gowda, M., Hatfield, J., Zhou, B., Mazur, E.,
 Kudrna, D., Dean, R., Soderlund, C., Wing, R. and Wang, G.
 TITLE Large-scale identification of ESTs involved in the interaction
 between rice and Magnaporthe grisea
 JOURNAL Unpublished
 COMMENT Contact: Rod Wing
 Arizona Genomics Institute
 University of Arizona
 Biological Sciences West, 489A, P.O. Box 210088, Tucson, AZ
 85721-0088, USA
 Tel: 520 626 3967
 Fax: 520 621 9288
 Email: http://genome.arizona.edu
 PCR Primers
 FORWARD: gta aaa cga cgg cca gtc
 BACKWARD: gga aac agc tat gac cat g
 Plate: 08 row: C column: 12
 Seq primer: gga aac agc tat gac cat g.
 Location/Qualifiers
 1..727
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="OSJNEC08C12"
 /tissue_type="Leaf"
 /dev_stage="3 week"
 /lab_host="DH10B"
 /clone_lib="OSJNEC"
 /note="Vector: pBluescript II KS+; Site 1: EcoRI; Site 2:
 XhoI; 6 hrs after inoculation with Rice Blast (C9240-1)"

BASE COUNT 204 a 192 c 166 g 165 t

ORIGIN
 Alignment Scores:
 Pred. No.: 4,18e-63 Length: 727
 Score: 744.00 Matches: 134
 Percent Similarity: 97.84% Conservative: 2
 Best Local Similarity: 96.40% Mismatches: 2
 Query Match: 92.65% Indels: 1
 DB: 14 Gaps: 0

US-09-868-025-2 (1-143) x CB655085 (1-727)

QY 1 MetValAspThrAsnProIleSerGluGlnSerGluThrHisAlaTrpCysTrpSer 20
 Db CTGTTAGATACAAATTTTCCAAATCAGTGAGCAATCAGAAACACATGCTTGGTGTGGAGC 281

QY 21 SerSerThrThrArgSerProSerArgHisLeuHisArgGluArgIleProCysLeu 40
 Db AGCAGACACACAGAGCCGCTCCGCCATCACTTCATCGAGCGGATCCCATGCTT 341

QY 41 AlaLeuGlyValThrAlaIleCysSerLeuValTrpIleHisValSerHisGlyGly 60
 Db GCCCTCGGTGTCACTGCAATTTCCAGTCTTGTGGATACATGTCAGCCACGGTGTGGA 401

QY 61 IleSerValAlaLeuCysSerGlnCysLeuGlnThrAsnAlaLeuArgProAsp 80
 Db ATCAGCGTGCCTTTTCAGCCAGTGCCTTCAGACGATGCGTCCGCCCTCGACCGAT 461

QY 81 CysLeuThrAsnAsnGlyGlyCysTyrglyCysHisGlySerLeuGlyHisValAsp 100
 Db TGTCTGACCAACAATGGAGGCTGCTATGCTGAATGCCATGCCAGTCTTGGCGCATGTGAC 521

QY 101 ArgPheProGlnHisSerAsnGluTrpAsnSerGlyMetAspSerCysValProLeuArg 120
 Db AGATTTCTCAGCATTCATATGAATGAACAGCGGCGATGACAGCTGCCAACCACTGAGG 581

QY 121 GlyCluPheLeuGlyValLeuThrProHisProLysMetGlu-PheAlaAlaIle 138
 Db 582 GGAGATTCTTGTGTGCTCACGCCACACCCCAAGAGGGATGTTTGTGTCATT 636

RESULT 3
 CB663774 763 bp mRNA linear EST 09-APR-2003
 LOCUS OSJNE09B20 x OSTNED Oryza sativa (japonica cultivar-group) cDNA
 DEFINITION clone OSJNE09B20 3', mRNA sequence.
 ACCESSION CB663774
 VERSION CB663774.1 GI:29667499
 KEYWORDS EST.
 SOURCE Oryza sativa (japonica cultivar-group)
 ORGANISM Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 1 (bases 1 to 763)
 Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,B.,
 Kudrna,D., Dean,R., Soderlund,C., Wang,R. and Wang,G.
 Large-scale identification of ESTs involved in the interaction
 between rice and Magnaporthe grisea
 Unpublished
 Contact: Rod Wing
 Arizona Genomics Institute
 University of Arizona
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
 85721-0088, USA
 Tel: 520 626 3967
 Fax: 520 621 9288
 Email: http://genome.arizona.edu
 PCR Primers
 FORWARD: gta aaa cga cgg cca gtc
 BACKWARD: gga aac agc tat gac cat g
 Plate: 09 row: B column: 20
 Seq primer: gga aac agc tat gac cat g.

FEATURES
 Source
 1..763
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="OSJNE09B20"
 /tissue_type="Leaf"
 /dev_stage="3 week"
 /lab_host="DH108"
 /clone_lib="OSJNEd"
 /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
 XhoI; 24 hrs after inoculation with Rice Blast (C9240-1)"
 BASE COUNT 217 a 202 c 177 g 167 t

Alignment Scores:
 Pred. No.: 4,488-63 Length: 763
 Score: 744.00 Matches: 134
 Percent Similarity: 97.84% Conservative: 2
 Best Local Similarity: 96.40% Mismatches: 2
 Query Match: 92.65% Indels: 1
 Gaps: 0

US-09-868-025-2 (1-143) x CB663774 (1-763)

QY 1 MetValAspThrAsnPheProIleSerGluInSerGluThrHisAlaTrpCysTrpSer 20
 Db 232 CTGGTAGATACAAATTTTCCCAATCAGTGAGCAATCAGAAACACATGCTTGGTGTGGAGC 291
 QY 21 SerSerThrThrArgSerProSerArgHisHisLeuHisArgGluArgIleProCysLeu 40
 Db 292 AGCAGCACCCAGAGCCCGCCGCCATCATCTTCGCGAGCGGATCCATGCTT 351
 QY 41 AlaLeuGlyValThrAlaIleCysSerLeuValTrpIleHisValSerHisGlyGly 60

Db 352 GCCCTCGGTGTCACAGCAATTTCCAGTCTTGTGTTGGATACATGTCAGCCACCGGGTGA 411
 QY 61 IleSerValAlaLeuCySerGlnCysLeuGlnThrAsnAlaLeuArgProArgProASP 80
 Db 412 ATCAGCGTGGCTCTTTTCAGCCAGTGGCTTTAGAGAAATGCGCTCGGCCCTGACCCGAT 471

QY 81 CysLeuThrAsnAsnGlyGlyCysTyrglyGluCysHisGlySerLeuGlyHisValasp 100
 Db 472 TGTCTGACCAACAATGGAGGCTGCTATGTTGAATGCAATGCGAGTCTTGGCATGTGTGAC 531
 QY 101 ArgPheProGlnHisSerAsnGluTrpAsnSerGlyMetAspSerCysLysProLeuArg 120
 Db 532 AGATTTCCTCAGCATTTCCAAATGAATGAACACGGGCGATGACAGCTGCAACACCTGAGG 591
 QY 121 GlyGluPheLeuGlyValLeuThrProHisProLysMetGlu-PheAlaAlaIle 138
 Db 592 GGAGATTCTTGTGTGCTCACGCCACACCCCAAGAGGGATGTTTGTGTCATT 646

RESULT 4
 CB000367/c 776 bp mRNA linear EST 10-JAN-2003
 LOCUS S345U H05 Rice cold stress germination cDNA library Oryza sativa
 DEFINITION (japonica cultivar-group) cDNA 5', mRNA sequence.
 ACCESSION CB000367
 VERSION CB000367.1 GI:27577672
 KEYWORDS EST.
 SOURCE Oryza sativa (japonica cultivar-group)
 ORGANISM Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 1 (bases 1 to 776)
 de los Reyes,B.G., Morsy,M., Gibbons,J., Varma,T.S.N., Antoine,W.,
 Redus,M., McGrath,J.W. and Halgren,R.
 Development of a chilling stress EST library of germinating rice
 (Oryza sativa L. subsp. japonica) enriched with stress-related and
 novel genes
 Unpublished
 Contact: Benildo G. de los Reyes
 Plant Genomics Lab., Department of Crop, Soil and Environmental
 Sciences
 University of Arkansas
 115 Plant Science Building, Fayetteville, AR 72701, USA
 Tel: (479)-575-7465
 Fax: (479)-575-8435
 Email: breyes@uark.edu
 Plate: S345U row: H column: 05
 Seq primer: T3
 Location/Qualifiers
 1..776
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="PI560247"
 /db_xref="taxon:39947"
 /tissue_type="coleotile, radicle, prophyll, immature leaf"
 /dev_stage="embryo at different stages of germination and
 shoots of germinated seeds under chilling stress
 (130C/100C)"
 /lab_host="SolR"
 /clone_lib="Rice cold stress germination cDNA library"
 /note="Organ: seedlings; Vector: Lambda Uni-Zap XR excised
 as pBluescript; Site 1: EcoRI; Site 2: XhoI; The cDNAs
 were derived from reverse transcription of mRNA samples
 from seeds at different stages of germination and
 seedlings at early phase of growth under chilling stress
 (130C/100C). The mRNA pool was used as template for double
 stranded cDNA synthesis using the Stratagene Uni-Zap XR
 cDNA synthesis and library kit. A total of 150,000 phages
 were excised from the primary library as pBluescript
 phagemid clones. Enrichment of the primary excised library
 with chilling-induced transcripts was performed by
 hybridizing the primary excised library colony lifts with
 the PCR-select subtraction product, with cold germinated

cDNA as tester and control temperature-germinated cDNA as driver."

BASE COUNT 170 a 180 c 203 g 223 t

ORIGIN

Alignment Scores:

Pred. No.: 4,59e-63 Length: 776
Score: 744.00 Matches: 134
Percent Similarity: 97.84% Conservations: 2
Best Local Similarity: 96.40% Mismatches: 2
Query Match: 92.65% Indels: 1
DB: 14 Gaps: 0

US-09-868-025-2 (1-143) x CB000367 (1-776)

QY 1 MetValAspThrAsnProIleSerGluGlnSerGluThrHisAlaTrpCysTrpSer 20
Db 533 CTGGTAGATACAAATTTCCAAATCAGTGGAGCAATCAGAACACATCTTGGTGGAGC 474
QY 21 SerSerThrThrArgSerProSerArgHisHisLeuHisArgGluArgLeuProCysLeu 40
Db 473 AGCAGCACCACCAAGAGCCGTCGGCCATCTTCTTCAGCGAGCGGATCCCATGCCCT 414
QY 41 AlaLeuGlyValThrAlaIleCysSerLeuValTrpIleHisValSerHisGlyGly 60
Db 413 GCCTCGGTCTACTGCAATTTCCAGTCTTGTGGATACATGTCAGCCACCGTGTGA 354
QY 61 IleSerValAlaLeuCysSerGlnCysLeuGlnThrAsnAlaLeuArgProArgProAsp 80
Db 353 ATCAGCGTCTCTTTTCAGCCAGTGGCTTCAGACGAATCGCTCGCCCTGACCCGAT 294
QY 81 CysLeuThrAsnAsnGlyGlyCysTrpGlyCysHisGlySerLeuGlyHisValAsp 100
Db 293 TGCTGACCAACCAATGGAGGCTGTATGTGTGAATGCCATGCCAGTCTTGGCATGTTGAC 234
QY 101 ArgPheProGlnHisSerArgGluTrpAsnSerGlyMetAspSerCysLysProLeuArg 120
Db 233 AGATTTCTCAGCATTCCTCAATGAATGAACAGCGGCATGGACAGCTGCAACACCTGAGG 174
QY 121 GlyGluPheLeuGlyValLeuThrProHisProLysMetGlu-PheAlaAlaIle 138
Db 173 GGAGATTTCTTGGTGTGCTCACGCCACACCCCAAGAGGATGTTGCTGCCAT 119

RESULT 5

CB627453 777 bp mRNA linear EST 08-APR-2003
LOCUS OSIIIEB02020.r OSIIIEB Oryza sativa (indica cultivar-group) cDNA
DEFINITION

clone OSIIIEB02020 3', mRNA sequence.

ACCESSION

CB627453

CB627453.1 GI:29622442

VERSION

EST.

KEYWORDS

Oryza sativa (indica cultivar-group)

SOURCE

Oryza sativa (indica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzaceae; Oryza.

1 (bases 1 to 777)

Jantasuriyarat,C., Lu G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,

Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.

Large-scale identification of ESTs involved in the interaction

between rice and Magnaporthe grisea

Unpublished

Contact: Rod Wing

Arizona Genomics Institute

University of Arizona

Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ

85721-0088, USA

Tel: 520 626 3967

Fax: 520 621 9288

Email: <http://genome.arizona.edu>

PCR Primers

FORWARD: gta aac cga cgg cca gtc

BACKWARD: gga aac agc tat gac cat g

Plate: 02 row: G column: 20
Seq primer: gga aac agc tat gac cat g.

FEATURES

source

1..777 /organism="Oryza sativa (indica cultivar-group)"

/mol_type="RNA"

/cultivar="IR36"

/db_xref="taxon:39946"

/clone="OSIIIEB02020"

/tissue_type="Leaf"

/dev_stage="3 week"

/lab_host="DH10B"

/clone_lib="OSIIIEB"

/note="vector: p Bluescript II KS +; Site 1: EcoRI; Site 2:

XhoI; 24 hrs after inoculation with Rice Blast (PO6-6-3)"

221 a 206 c 173 g 177 t

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.: 4,59e-63 Length: 777
Score: 744.00 Matches: 134
Percent Similarity: 97.84% Conservations: 2
Best Local Similarity: 96.40% Mismatches: 2
Query Match: 92.65% Indels: 1
DB: 14 Gaps: 0

US-09-868-025-2 (1-143) x CB627453 (1-777)

QY 1 MetValAspThrAsnProIleSerGluGlnSerGluThrHisAlaTrpCysTrpSer 20
Db 240 CTGGTAGATACAAATTTCCAAATCAGTGGAGCAATCAGAACACATCTTGGTGGAGC 299
QY 21 SerSerThrThrArgSerProSerArgHisHisLeuHisArgGluArgLeuProCysLeu 40
Db 300 AGCAGCACCACCAAGAGCCGTCGGCCATCTTTCAGCGAGCGGATCCCATGCCCT 359
QY 41 AlaLeuGlyValThrAlaIleCysSerLeuValTrpIleHisValSerHisGlyGly 60
Db 360 GCCTCGGTCTACTGCAATTTCCAGTCTTGTGGATACATGTCAGCCACCGTGTGA 419
QY 61 IleSerValAlaLeuCysSerGlnCysLeuGlnThrAsnAlaLeuArgProArgProAsp 80
Db 420 ATCAGCGTCTCTTTTCAGCCAGTGGCTTCAGACGAATCGCTCGCCCTCGACCCGAT 479
QY 81 CysLeuThrAsnAsnGlyGlyCysTrpGlyCysHisGlySerLeuGlyHisValAsp 100
Db 480 TGTCTGACCAACAAATGGAGGCTGTATGTGAATGCCATGGCAGTCTTGGCATGTTGAC 539
QY 101 ArgPheProGlnHisSerArgGluTrpAsnSerGlyMetAspSerCysLysProLeuArg 120
Db 540 AGATTTCTCAGCATTCCTCAATGAATGAACAGCGGCATGGACAGCTGCAACACCTGAGG 599
QY 121 GlyGluPheLeuGlyValLeuThrProHisProLysMetGlu-PheAlaAlaIle 138
Db 600 GGAGATTTCTTGGTGTGCTCACGCCACACCCCAAGAGGATGTTTGTGCCAT 654

RESULT 6

CB655151

CB655151

LOCUS

DEFINITION

clone OSJNEC08E12 3', mRNA sequence.

CB655151

CB655151.1 GI:29658876

EST.

KEYWORDS

Oryza sativa (japonica cultivar-group)

Oryza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzaceae; Oryza.

1 (bases 1 to 780)

Jantasuriyarat,C., Lu G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,

Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.

Large-scale identification of ESTs involved in the interaction

between rice and Magnaporthe grisea

Unpublished

Contact: Rod Wing

Arizona Genomics Institute

University of Arizona

Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ

85721-0088, USA

Tel: 520 626 3967

Fax: 520 621 9288

between rice and Magnaporthe grisea
 Unpublished
 Contact: Rod Wing
 Arizona Genomics Institute
 University of Arizona
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
 85721-0088, USA
 Tel: 520 626 3967
 Fax: 520 621 9288
 Email: http://genome.arizona.edu
 PCR Primers
 FORWARD: gta aaa cga cgg cca gtc
 BACKWARD: gga aac agc tat gac cat g
 Plate: 08 row: E column: 12
 Seq primer: gga aac agc tat gac cat g
 Location/Qualifiers
 1. 780
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="OSJNEC08E12"
 /tissue_type="Leaf"
 /dev_stage="3 week"
 /lab_host="DH10B"
 /clone_lib="OSJNEC"
 /note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2: XhoI; 6 hrs after inoculation with Rice Blast (C9240-1)"
 XhoI: 206 c 176 g 185 t

BASE COUNT 213 a 206 c 176 g 185 t
 ORIGIN

Alignment Scores:
 Pred. No.: 4.62e-63 Length: 780
 Score: 744.00 Matches: 134
 Percent Similarity: 97.84% Conservative: 2
 Best Local Similarity: 96.40% Mismatches: 2
 Query Match: 92.65% Indels: 1
 DB: 14 Gaps: 0

US-09-868-025-2 (1-143) x CB655151 (1-780)

Qy 1 MetValAspThrAsnProIleSerGluGlnSerGluThrHisAlaTrpCysTrpSer 20
 Db 280 CTGTAGATACAAATTTTCCAAATCAGTGAGCAATCAGAAACACATGCTTGGTGTGGAGC 339

Qy 21 SerSerThrThrArgSerProSerArgHisLeuHisArgGluArgIleProCysLeu 40
 Db 340 AGCAGCACCCAGAGCCGTCGCCCATCATCTTCATCGAGCGGATCCCATGCGCTT 399

Qy 41 AlaLeuGlyValThrAlaIleCysSerLeuValTrpIleHisValSerHisGlyGly 60
 Db 400 GCCCTCGGTGTCACGTGCAATTTCCAGTCTTTGGATACATGTCACCCACCGGTGGG 459

Qy 61 IleSerValAlaLeuCysSerGlnCysLeuGlnThrAsnAlaLeuArgProArgProAsp 80
 Db 460 ATCAGCGTCGCTCTTTCAGCGCAGTCGCTTCAGACGAATGCGCTCCGCCCTCGACCGAT 519

Qy 81 CysLeuThrAsnArgGlyCysTrpGlyCysHisGlySerLeuGlyHisValAsp 100
 Db 520 TGTCTGACCAACAATGAGGCTGTATGTAATGCCATGCGAGTCTTGGCGCATGTGAC 579

Qy 101 ArgPheProGlnHisSerAsnGluTrpAsnSerGlyMetAspSerCysLysProLeuArg 120
 Db 580 AGATTCTCAGCATTCCAATGAATGAACAGCGCATGACAGCTGCMAACCACTGAGG 639

Qy 121 GlyGluPheLeuGlyValLeuThrProHisProLysMetGlu-PheAlaIle 138
 Db 640 GGAGAAATTTCTGTGTGCTCACGCCACACCCCAAGAGGATGTTTGTGCCATT 694

RESULT 7
 LOCUS CB647260 792 bp mRNA linear EST 08-APR-2003
 DEFINITION OSJNEB10B21.1 OSJNEB Oryza sativa (japonica cultivar-group) cDNA

clone OSJNEB10B21.3', mRNA sequence.
 CB647260
 VERSION CB647260.1 GI:29642253
 EST
 SOURCE Oryza sativa (japonica cultivar-group)
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.
 1 (bases 1 to 792)
 Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E., Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
 Large-scale identification of ESTs involved in the interaction between rice and Magnaporthe grisea
 Unpublished
 Contact: Rod Wing
 Arizona Genomics Institute
 University of Arizona
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
 85721-0088, USA
 Tel: 520 626 3967
 Fax: 520 621 9288
 Email: http://genome.arizona.edu
 PCR Primers
 FORWARD: gta aaa cga cgg cca gtc
 BACKWARD: gga aac agc tat gac cat g
 Plate: 10 row: B column: 21
 Seq primer: gga aac agc tat gac cat g
 Location/Qualifiers
 1. 792
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="OSJNEB10B21"
 /tissue_type="Leaf"
 /dev_stage="3 week"
 /lab_host="DH10B"
 /clone_lib="OSJNEB"
 /note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2: XhoI; 24 hrs after inoculation with Rice Blast (Che 86061)"
 XhoI: 222 a 212 c 184 g 174 t

BASE COUNT 222 a 212 c 184 g 174 t
 ORIGIN

Alignment Scores:
 Pred. No.: 4.72e-63 Length: 792
 Score: 744.00 Matches: 134
 Percent Similarity: 97.84% Conservative: 2
 Best Local Similarity: 96.40% Mismatches: 2
 Query Match: 92.65% Indels: 1
 DB: 14 Gaps: 0

US-09-868-025-2 (1-143) x CB647260 (1-792)

Qy 1 MetValAspThrAsnProIleSerGluGlnSerGluThrHisAlaTrpCysTrpSer 20
 Db 233 CTGTAGATACAAATTTTCCAAATCAGTGAGCAATCAGAAACACATGCTTGGTGTGGAGC 292

Qy 21 SerSerThrThrArgSerProSerArgHisLeuHisArgGluArgIleProCysLeu 40
 Db 293 AGCAGCACCCAGAGCCGTCGCCCATCATCTTCATCGAGCGGATCCCATGCGCTT 352

Qy 41 AlaLeuGlyValThrAlaIleCysSerLeuValTrpIleHisValSerHisGlyGly 60
 Db 353 GCCCTCGGTGTCACGTGCAATTTCCAGTCTTTGGATACATGTCACCCACCGGTGGG 412

Qy 61 IleSerValAlaLeuCysSerGlnCysLeuGlnThrAsnAlaLeuArgProArgProAsp 80
 Db 413 ATCAGCGTCGCTCTTTCAGCGCAGTCGCTTCAGACGAATGCGCTCCGCCCTCGACCGAT 472

Qy 81 CysLeuThrAsnArgGlyCysTrpGlyCysHisGlySerLeuGlyHisValAsp 100

DB 473 TGTCTGACCAACAATGAGCGCTGCTATGGTGAATGCCATGGCAGTCCTTTGGGCATGTTGAC 532

QY 101 ArgPheProGlnHisSerAsnGluTrpAanSerGlyMetAspSerCysLeuProLeuArg 120
533 AGATTCTCTCAGCATTCCTCAATGAATGAACAGCGGCATGCACAGCTGCAAAACCACCTGAGG 592

DB 121 GlyGluPheLeuGlyValLeuThrProHisProLysMetGlu-PheAlaAlaIle 138
QY : : : : :
DB 593 GGAGAATTTCTTGTTGTTGCTCAGCCACACCCCAGAGGGAGTTTGTGTCANT 647

RESULT 8
EB2620522
LOCUS
DDBFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CB620522 799 bp mRNA linear EST 08-APR-2003
OSIIIEA05M20.r OSIIIEA Oryza sativa (indica cultivar-group) cDNA
clone OSIIIEA05M20 3', mRNA sequence.
CB620522.1 GI:29615510
EST.
Oryza sativa (indica cultivar-group)
Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophytes; Magnoliophyta; Liliopsida; Poales; Poaceae;
Erbartoideae; Oryzaceae; Oryza.
1 (bases 1 to 799)
Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
Unpublished
Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR Primers
FORWARD: gta aaa cga cgg cca gat g
BACKWARD: gga aac agc tat gac cat g
Plate: 05 row: M column: 20
Seq primer: gga aac agc tat gac cat g.

FEATURES
source
location/Qualifiers
1..799
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"
/cultivar="IR36"
/db_xref="taxon:39946"
/clone="OSIIIEA05M20"
/tissue_type="Leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="OSIIIEA"
/note="Vector:pBluescript II KS +; Site_1: EcoRI; Site_2:
XhoI; Lesion Mimic SPL II"

BASE COUNT 207 a 220 c 194 g 178 t
ORIGIN

Alignment Scores:

Pred. No.:	4,78e-63	Length:	799
Score:	744.00	Matches:	134
Percent Similarity:	97.84%	Conservative:	2
Best Local Similarity:	96.40%	Mismatches:	1
Query Match:	92.65%	Indels:	1
DB:	14	Gaps:	0

US-09-868-025-2 (1-143) x CB620522 (1-799)

QY 1 MetValAsnThrAsnPheProLeuSerGlnHisSerGluThrHisAlaTrpCysTrpSer 20
: : : : :
DB 153 CTGTGTAGATACAAATTTTCCAATCATGTGAGCAATCAGAACACATGCTTGTTGGAGC 212

QY 21 SetSerThrArgSerProSerArgHisHisLeuHisArgGluArgGileProCysLeu 40

```

Score: 744.00 Matches: 134
Percent Similarity: 97.84% Conservative: 2
Best Local Similarity: 96.40% Mismatches: 2
Query Match: 92.65% Indels: 1
DB: 14 Gaps: 0

US-09-868-025-2 (1-143) x CB660999 (1-805)

QY 1 MetValAspThrAsnPheProIleSerGluGlnSerGluThrHisAlaTrpCysTrpSer 20
DB 232 CTGGTAGATACAAATTTTCCAAATCAGTGAACATCAGAACACATGCTTGGTGGAGC 291
QY 21 SerSerThrThrArgSerProSerArgHisHisLeuHisArgGluArgIleProCysLeu 40
DB 292 AGCAGCACCACCAAGAGCCGTCGCCCATCATCTTCATCGCAGCGGATGCCATGCCCT 351
QY 41 AlaLeuGlyValThrAlaIleCysSerLeuValTrpIleHisValSerHisGlyGly 60
DB 352 GCCTCGGTGTCTGCTGCAATTTCCAGTCTTGTGGATATGTCACCCACGCTGGTGA 411
QY 61 IleSerValAlaLeuCysSerGlnCysLeuGlnThrAsnAlaLeuArgProArgProAsp 80
DB 412 ATCAGCGTGTCTTTTCAGCCAGTGTGCTTCAGAGCAATGCGTCCGCCCTCGACCCGAT 471
QY 81 CysLeuThrAsnGlyGlyCysTrpGlyGlyCysHisGlySerLeuGlyHisValAsp 100
DB 472 TGTCTGACCAACAAATGAGGCTGTATGTGTAATGCCATGCGCATCTTGGCATGTGAC 531
QY 101 ArgPheProGlnHisSerAsnGluTrpAsnSerGlyMetAspSerCysLysProLeuArg 120
DB 532 AGATTCTCAGCATTCATGAATGGAACAGCGGATGACAGCAGCTGCAACCACTGAGG 591
QY 121 GlyGluPheLeuGlyValLeuThrProHisProLysMetGlu-PheAlaAlaIle 138
DB 592 GGAGAAATTTCTTGTGTGCTCACGCCACACCCCAAGAGGATGTTTGTGCCATT 646

RESULT 10
CB633818 816 bp mRNA linear EST 08-APR-2003
LOCUS OSJNEC12016.1 OSJNEC Oryza sativa (indica cultivar-group) cDNA
clone OSJNEC12016.3, mRNA sequence.
CB633818
CB633818.1 GI:29628807
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Oryza sativa (indica cultivar-group)
Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE
1 (bases 1 to 816)
Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
Unpublished
Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR Primers
FORWARD: gta aaa cga cgg cca gat g
BACKWARD: gga aac agc tat gac cat g
Plate: 12 row: 0 column: 16
Seq primer: gga aac agc tat gac cat g.
Location/Qualifiers
1. .816
/organism="Oryza sativa (indica cultivar-group)"
/mol type="mRNA"
/cultivar="IR36"

FEATURES
source
/db_xref="taxon:39946"
/clone="OSJNEC12016"
/tissue type="Leaf"
/dev stage="3 week"
/lab_hosts="DH10B"
/clone_lib="OSJNEC"
/note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:
XhoI; 24 hrs after inoculation with Rice Blast (PO6-6-3)"
BASE COUNT 214 a 222 c 193 g 187 t
ORIGIN
Alignment Scores:
Pred. No.: 4,938-63 Length: 816
Score: 744.00 Matches: 134
Percent Similarity: 97.84% Conservative: 2
Best Local Similarity: 96.40% Mismatches: 2
Query Match: 92.65% Indels: 1
DB: 14 Gaps: 0
US-09-868-025-2 (1-143) x CB633818 (1-816)
QY 1 MetValAspThrAsnPheProIleSerGluGlnSerGluThrHisAlaTrpCysTrpSer 20
DB 184 CTGGTAGATACAAATTTTCCAAATCAGTGAACATCAGAACACATGCTTGGTGGAGC 243
QY 21 SerSerThrThrArgSerProSerArgHisHisLeuHisArgGluArgIleProCysLeu 40
DB 244 AGCAGCACCACCAAGAGCCGTCGCCCATCATCTTCATCGCAGCGGATGCCATGCCCT 303
QY 41 AlaLeuGlyValThrAlaIleCysSerLeuValTrpIleHisValSerHisGlyGly 60
DB 304 GCCTCGGTGTCTGCTGCAATTTCCAGTCTTGTGGATATGTCACCCACGCTGGTGA 363
QY 61 IleSerValAlaLeuCysSerGlnCysLeuGlnThrAsnAlaLeuArgProArgProAsp 80
DB 364 ATCAGCGTGTCTTTTCAGCCAGTGTGCTTCAGAGCAATGCGTCCGCCCTCGACCCGAT 423
QY 81 CysLeuThrAsnGlyGlyCysTrpGlyGlyCysHisGlySerLeuGlyHisValAsp 100
DB 424 TGTCTGACCAACAAATGAGGCTGTATGTGTAATGCCATGCGCATCTTGGCATGTGAC 483
QY 101 ArgPheProGlnHisSerAsnGluTrpAsnSerGlyMetAspSerCysLysProLeuArg 120
DB 484 AGATTCTCAGCATTCATGAATGGAACAGCGGATGACAGCAGCTGCAACCACTGAGG 543
QY 121 GlyGluPheLeuGlyValLeuThrProHisProLysMetGlu-PheAlaAlaIle 138
DB 544 GGAGAAATTTCTTGTGTGCTCACGCCACACCCCAAGAGGATGTTTGTGCCATT 598

RESULT 11
CB659692 836 bp mRNA linear EST 09-APR-2003
LOCUS OSJNEC16006.1 OSJNEC Oryza sativa (japonica cultivar-group) cDNA
clone OSJNEC16006.3, mRNA sequence.
CB659692
CB659692.1 GI:29663417
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE
1 (bases 1 to 836)
Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
Unpublished
Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA

```

Tel: 520 626 3967
 Fax: 520 621 9288
 Email: http://genome.arizona.edu

PCR Primers
 FORWARD: gta aac cga cgg cca gtc
 BACKWARD: gga aac agc tat gac cat g
 Plate: 16 row: 0 column: 06
 Seq primer: gga aac agc tat gac cat g.

FEATURES
 source
 1..836
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="OSJNB16006"
 /tissue_type="Leaf"
 /dev_stage="3 week"
 /lab_host="DH10B"
 /clone_lib="OSJNEC"
 /note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2: XhoI; 6 hrs after inoculation with Rice_Blast (C9240-1)"
 BASE COUNT 208 a 227 c 201 g 200 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 5,1e-63 Length: 836
 Score: 744.00 Matches: 134
 Percent Similarity: 97.84% Conservative: 2
 Best Local Similarity: 96.40% Mismatches: 2
 Query Match: 92.65% Indels: 1
 DB: 14 Gaps: 0

US-09-868-025-2 (1-143) x CB619892 (1-836)

QY 1 MetValAspThrAsnProIleSerGluGlnSerGluThrHisAlaTrpCysTrpSer 20
 Db 153 CTGGTAGATACAAATTTTCCAAATCAGTGAGCAATCAGAAACACATGCTTGGTGTGGAGC 212

QY 21 SerSerThrThrArgSerProSerArgHisHisLeuHisArgGluArgIleProCysLeu 40
 Db 213 AGCAGCACCAACAGAGCCGCTCCGCCATCATCTTCATCGAGCGGATCCCATGCCIT 272

QY 41 AlaLeuGlyValThrAlaIleCysSerLeuValTrpIleHisValSerHisGlyGly 60
 Db 273 GCCCTCGGTCTCACTGCAATTTCCAGTCTTGTGGATACATGTCAGCCACGGTGTGGA 332

QY 61 IleSerValAlaLeuCysSerGlnCysLeuGlnThrAsnAlaLeuArgProArgProAsp 80
 Db 333 ATCAGCGCTGCTCTTTGACGACCACTGCTTCAGACGAATGGCTCCGCCCTCGACCGAT 392

QY 81 CysLeuThrAsnAsnGlyGlyCysTrpGlyGluCysHisGlySerLeuGlyHisValAsp 100
 Db 393 TGTCTGACCAACAAATGAGGCTGCTATGGTGAATGCCATGCGCAGTCTTGGCCATGTGAC 452

QY 101 ArgPheProGlnHisSerAsnGluTrpAsnSerGlyMetAspSerCysLysProLeuArg 120
 Db 453 AGATTTCTTCAGCATTCCTCAATGAATGAACAGCGGCATGACAGCTGCAAAACCACTGAGG 512

QY 121 GlyGluPheLeuGlyValLeuThrProHisProLysMetGlu-PheAlaAlaIle 138
 Db 513 GGAGAAATTTCTTGTGTGCTACGCCACACCCCAAGAGGATGTTGCTGCCATT 567

Tel: 520 626 3967
 Fax: 520 621 9288
 Email: http://genome.arizona.edu

PCR Primers
 FORWARD: gta aac cga cgg cca gtc
 BACKWARD: gga aac agc tat gac cat g
 Plate: 16 row: 0 column: 06
 Seq primer: gga aac agc tat gac cat g.

FEATURES
 source
 1..836
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="OSJNB16006"
 /tissue_type="Leaf"
 /dev_stage="3 week"
 /lab_host="DH10B"
 /clone_lib="OSJNEC"
 /note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2: XhoI; 6 hrs after inoculation with Rice_Blast (C9240-1)"
 BASE COUNT 208 a 227 c 201 g 200 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 5,1e-63 Length: 836
 Score: 744.00 Matches: 134
 Percent Similarity: 97.84% Conservative: 2
 Best Local Similarity: 96.40% Mismatches: 2
 Query Match: 92.65% Indels: 1
 DB: 14 Gaps: 0

US-09-868-025-2 (1-143) x CB619892 (1-836)

QY 1 MetValAspThrAsnProIleSerGluGlnSerGluThrHisAlaTrpCysTrpSer 20
 Db 153 CTGGTAGATACAAATTTTCCAAATCAGTGAGCAATCAGAAACACATGCTTGGTGTGGAGC 212

QY 21 SerSerThrThrArgSerProSerArgHisHisLeuHisArgGluArgIleProCysLeu 40
 Db 213 AGCAGCACCAACAGAGCCGCTCCGCCATCATCTTCATCGAGCGGATCCCATGCCIT 272

QY 41 AlaLeuGlyValThrAlaIleCysSerLeuValTrpIleHisValSerHisGlyGly 60
 Db 273 GCCCTCGGTCTCACTGCAATTTCCAGTCTTGTGGATACATGTCAGCCACGGTGTGGA 332

QY 61 IleSerValAlaLeuCysSerGlnCysLeuGlnThrAsnAlaLeuArgProArgProAsp 80
 Db 333 ATCAGCGCTGCTCTTTGACGACCACTGCTTCAGACGAATGGCTCCGCCCTCGACCGAT 392

QY 81 CysLeuThrAsnAsnGlyGlyCysTrpGlyGluCysHisGlySerLeuGlyHisValAsp 100
 Db 393 TGTCTGACCAACAAATGAGGCTGCTATGGTGAATGCCATGCGCAGTCTTGGCCATGTGAC 452

QY 101 ArgPheProGlnHisSerAsnGluTrpAsnSerGlyMetAspSerCysLysProLeuArg 120
 Db 453 AGATTTCTTCAGCATTCCTCAATGAATGAACAGCGGCATGACAGCTGCAAAACCACTGAGG 512

QY 121 GlyGluPheLeuGlyValLeuThrProHisProLysMetGlu-PheAlaAlaIle 138
 Db 513 GGAGAAATTTCTTGTGTGCTACGCCACACCCCAAGAGGATGTTGCTGCCATT 567

RESULT 12
 CB619892
 LOCUS
 DEFINITION OSJNB04M03.2 OSJNBa Oryza sativa (indica cultivar-group) cDNA
 clone OSJNB04M03.3', mRNA sequence.
 ACCESSION CB619892
 VERSION CB619892.1 GI:29614879
 KEYWORDS EST.
 SOURCE Oryza sativa (indica cultivar-group)
 ORGANISM Oryza sativa (indica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Db 529 GGAGATTTCTTGCTGTGCTCAGCCACACCCCAAGAGGATGTTTGTGTCATT 583

RESULT 13
 CB647623
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

CB647623 682 bp mRNA linear EST 08-APR-2003
 OSJNEB10K01.r OSJNEB Oryza sativa (japonica cultivar-group) cDNA
 clone OSJNEB10K01 3', mRNA sequence.

CB647623
 CB647623.1 GI:29642616
 EST.

Oryza sativa (japonica cultivar-group)
 Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.

1 (bases 1 to 682)
 Jantasuriyarat, C., Lu G., Gowda, M., Hatfield, J., Zhou, B., Mazur, E.,
 Kudrna, D., Dean, R., Soderlund, C., Wing, R., and Wang, G.
 Large-scale identification of ESTs involved in the interaction
 between rice and Magnaporthe grisea
 Unpublished
 Contact: Rod Wing
 Arizona Genomics Institute
 University of Arizona
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
 85721-0088, USA
 Tel: 520 621 9288
 Fax: 520 621 9288
 Email: http://genome.arizona.edu

PCR Primers
 FORWARD: gta aaa cga cgg cca gtc
 BACKWARD: gga aac agc tat gac cat g
 Plate: 10 row: K column: 01
 Seq primer: gga aac agc tat gac cat g.

FEATURES
 source
 1..682
 Location/Qualifiers
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="OSJNEB10K01"
 /tissue_type="Leaf"
 /dev_stage="3 week"
 /lab_host="DH10B"
 /clone_lib="OSJNEB"
 /notes="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:
 XhoI; 24 hrs after inoculation with Rice Blast (Che 86061
)")

BASE COUNT 199 a 179 c 145 g 159 t

Alignment Scores:
 Pred. No.: 9,44e-63 Length: 682
 Score: 740.00 Matches: 133
 Percent Similarity: 97.84% Conservatives: 3
 Best Local Similarity: 95.68% Mismatches: 2
 Query Match: 92.15% Indels: 1
 DB: 14 Gaps: 0

US-09-858-025-2 (1-143) x CB647623 (1-682)

Qy 1 MetValAspThrAsnProIleSerGluGlnSerGluThrHisAlaTrpOxyTrpSer 20
 Db 238 CTGGTAGATACAAATTTTCCAACTAGTGAGCATCAAAACACATCTTGTTGGTGGAGC 297

Qy 21 SerSerThrArgSerProSerArgHisHisLeuHisArgGluArgIleProCysLeu 40
 Db 298 AGCAGCACCCAGAGCCCGTCCGCCATCATCTTCATCGAGCGGATCCCATGCGCTT 357

Qy 41 AlaLeuGlyValThrAlaIleCysSerLeuValTrpIleHisValSerHisGlyGly 60
 Db 358 GCCTCGGTGTCACTGCAATTTCCAGTCTTGTGGATACATGTCCAGCCAGCGTGGTGA 417

Qy 61 IleSerValAlaLeuCysSerGlnCysLeuGlnThrAsnAlaLeuArgProArgProAsp 80
 Db 418 ATCAGCGTGGCTCTTTGAGCCAGTGGCTTTCAGAGCAATGGCTCGGCCCTCGACCCGAT 477

Qy 81 CysLeuThrAsnAsnGlyGlyCysTrpGlyGluCysHisGlySerLeuGlyHisValAsp 100
 Db 478 TGTCTGACCAACAATGGAGGCTGTATGTAATGCCATGCGAGTCTTGGGCATGTTGAC 537

Qy 101 ArgPheProGlnHisSerAsnGluTrpAsnSerGlyMetAspSerCysLeuValProLeuArg 120
 Db 538 AGATTTCTCAGCATTCCTCAATGAATGGAACAGCGCATGGACAGCTGCAACCACTGAGG 597

Qy 121 GlyGluPheLeuGlyValLeuThrProHisProLysMetGlu-PheAlaAlaIle 138
 Db 598 GGAGAAATTTCTTGTTGCTCAGCCACACCCCAAGAGGATGTTTGTGTCATT 652

RESULT 14
 CB643278
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

CB643278 653 bp mRNA linear EST 08-APR-2003
 OSJNEB03002.r OSJNEB Oryza sativa (japonica cultivar-group) cDNA
 clone OSJNEB03002 3', mRNA sequence.

CB643278
 CB643278.1 GI:29638269
 EST.

Oryza sativa (japonica cultivar-group)
 Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.

1 (bases 1 to 653)
 Jantasuriyarat, C., Lu, G., Gowda, M., Hatfield, J., Zhou, B., Mazur, E.,
 Kudrna, D., Dean, R., Soderlund, C., Wing, R., and Wang, G.
 Large-scale identification of ESTs involved in the interaction
 between rice and Magnaporthe grisea
 Unpublished
 Contact: Rod Wing
 Arizona Genomics Institute
 University of Arizona
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
 85721-0088, USA
 Tel: 520 626 3967
 Fax: 520 621 9288
 Email: http://genome.arizona.edu

PCR Primers
 FORWARD: gta aaa cga cgg cca gtc
 BACKWARD: gga aac agc tat gac cat g
 Plate: 03 row: O column: 02
 Seq primer: gga aac agc tat gac cat g.

FEATURES
 source
 1..653
 Location/Qualifiers
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="OSJNEB03002"
 /tissue_type="Leaf"
 /dev_stage="3 week"
 /lab_host="DH10B"
 /clone_lib="OSJNEB"
 /notes="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:
 XhoI; 24 hrs after inoculation with Rice Blast (Che 86061
)")

BASE COUNT 168 a 177 c 149 g 159 t

Alignment Scores:
 Pred. No.: 1.75e-62 Length: 653
 Score: 737.00 Matches: 132
 Percent Similarity: 97.84% Conservatives: 4
 Best Local Similarity: 94.96% Mismatches: 2
 Query Match: 91.78% Indels: 1
 DB: 14 Gaps: 0

Best Local Similarity: 95.71%

```

DB: 9 Gaps: 0
US-09-868-025-2 (1-143) x AU075373 (1-712)
QY 1 MetValAspThrAsnPhenProIleSerGluGlnSerGluThrHisalaTrpCysTrpSer 20
Db 465 CTGGTAGATACAAATTTTCCAATCAGTGAGCAATCAGAAACACATGCTTGGTGTTCGAGC 406
QY 21 SerSerThrThrArgSerProSerArgHisHisIleuHisArgGluuargIleProCysLeu 400
Db 405 AGCAGCACCCACAGAACGCCGCTCCGCGCATCATCTTCATCGCAGGCGGATCCCATGCCCT 346
QY 41 AlaLeuGlyValThrAlaIleCysSerLeuValTrpIleHisValSerHisGlyGlyGly 60
Db 345 GCCCTCGGTGTCACTGCATNTTCCAGTCTTGTGTGGATACATGTCAGCCACGGTGTGTGA 286
QY 61 IleSerValAlaLeuCysSerGlnCysLeuGlnThrAsnAlaLeuArgProArgProAsp 80
Db 285 ATCAGCGTCGCTCTTTGCAGCCAGTCGCTTCAGACGAAATGCGCTCCGCGCTCGACCCGAT 226
QY 81 CysLeuThrAsnAsnGlyGlyCysGlyGlyGluCysHisGlySerLeuGlyHisValAsp 100
Db 225 TGCTCGACCAACAATGGAGGGTGTGTATGGTGAATGCCATGCCAGTCTTGGCGCATGTGAC 166
QY 101 ArgPheProGlnHisSerAsnGluTrpAsnSerGlyMetAspSerCysIleYsProLeuArg 120
Db 165 AGATTTTCTCAGCATTTCCATGAATGAAACAGCGGCATGCACAGCTGCAAAACCATGAGG 106
QY 121 GlyGlu-PheLeuGlyValLeuThrProHisIleProIysMetGlu-PheAlaAlaIle 138
Db 105 CGAGAAATTTCTTGTTGTGTGCTCAGCCACACCCCAAGAGGATGTTTGTCTGCCATT 50

```

Search completed: November 14, 2003, 13:41:05
Job time : 1919 secs

QY 121 GlyGlu-PheLeuGlyValLeuThrProHisProLysMetGlu-PheAlaAlaLeu 138
DB 105 GGAGAAATTTCTTGGTGTGCTACGCCACACCCCAAGAGGGATGTTTGTGCCATT 50
:::

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 13:25:15 ; Search time 35 Seconds
(without alignments)
1054.329 Million cell updates/sec

Title: US-09-868-025-2
Perfect score: 143
Sequence: 1 MVDNFPISQSTHAWCWS.....LGLTTPHPRKMEFAAIRAGKV 143

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_23.:

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_ivirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	143	100.0	143	10 Q9SN25	Q9sn25 oryza sativ
2	8	5.6	271	5 Q9V8B2	Q9v8b2 drosophila
3	8	5.6	295	16 Q9YU3	Q9yu3 anabena sp
4	8	5.6	623	10 P3J484	P3j484 pisum sativ
5	8	5.6	624	10 Q48662	Q48662 cucurbita c
6	8	5.6	1431	11 Q8CFM6	Q8cfm6 rattus norv
7	8	5.6	2559	11 Q9R4U0	Q9r4u0 mus musculu
8	7	4.9	64	2 Q04390	Q04390 streptomyc
9	7	4.9	124	16 Q8XPH0	Q8xph0 ralsstonia s
10	7	4.9	174	16 Q9RCW2	Q9rcw2 streptomyc
11	7	4.9	185	9 Q9XJD9	Q9xjd9 streptococc
12	7	4.9	204	2 Q8KXV5	Q8kxv5 escherichia
13	7	4.9	207	10 Q9LMS9	Q9lms9 arabidopsis
14	7	4.9	243	16 Q92VP1	Q92vp1 rhizobium m
15	7	4.9	244	12 Q98803	Q98803 tobacco bus
16	7	4.9	249	5 Q9GUJ2	Q9guj2 caenorhabdi

17	7	4.9	249	12 Q9ILZ8	Q9ilz8 groundnut r
18	7	4.9	249	12 Q9IM01	Q9im01 groundnut r
19	7	4.9	249	12 Q9ILZ9	Q9ilz9 groundnut r
20	7	4.9	249	12 Q9ILZ7	Q9ilz7 groundnut r
21	7	4.9	249	12 Q9ILZ6	Q9ilz6 groundnut r
22	7	4.9	249	12 Q9IM00	Q9im00 groundnut r
23	7	4.9	249	12 Q9ILZ4	Q9ilz4 groundnut r
24	7	4.9	249	12 Q9ILZ5	Q9ilz5 groundnut r
25	7	4.9	260	12 Q67685	Q67685 groundnut r
26	7	4.9	274	16 Q988N5	Q988n5 rhizobium l
27	7	4.9	280	16 Q9KPV7	Q9kpv7 vibrio chol
28	7	4.9	283	16 Q8E5S8	Q8e5s8 streptococc
29	7	4.9	283	16 Q8E0S1	Q8e0s1 streptococc
30	7	4.9	303	16 Q8U4Z3	Q8u4z3 agrobacteri
31	7	4.9	318	16 Q92T15	Q92t15 rhizobium m
32	7	4.9	331	16 Q986M6	Q986m6 rhizobium m
33	7	4.9	333	16 Q92K23	Q92k23 rhizobium m
34	7	4.9	354	16 Q8R8Z8	Q8r8z8 thermocanaer
35	7	4.9	377	16 Q8ZK44	Q8zk44 salmonella
36	7	4.9	377	16 Q8Z142	Q8z142 salmonella
37	7	4.9	386	16 P96838	P96838 mycobacteri
38	7	4.9	404	16 Q8K7Y4	Q8k7y4 streptococc
39	7	4.9	415	16 Q9KKX9	Q9kkx9 streptomyce
40	7	4.9	419	16 Q9A0C9	Q9a0c9 streptococc
41	7	4.9	419	16 Q8P1C2	Q8p1c2 streptococc
42	7	4.9	429	10 Q8LFPK3	Q8lfpk3 arabidopsis
43	7	4.9	432	11 Q8K0K2	Q8k0k2 mus musculu
44	7	4.9	432	11 Q9P63	Q9p63 rattus norv
45	7	4.9	432	11 Q8CD10	Q8cd10 mus musculu

ALIGNMENTS

RESULT 1
Q9SN25 PRELIMINARY; PRT; 143 AA.
AC Q9SN25; 01-MAY-2000 (TREMREL. 13, Created)
DT 01-MAY-2000 (TREMREL. 13, Last sequence update)
DT 01-OCT-2002 (TREMREL. 22, Last annotation update)
DE Hypothetical 15.7 kDa protein (Fragment).
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. IR64; TISSUE=Seedling;
RA Patell V.M., Mathai C.A., Divya C., Ashok M.;
RT "Oryza sativa Variety IR64(CDNA clone AGTSAL-11 from 7 days old seedling)".
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF192975; AAF06789.1; -
KW Gramene; Q9SN25; -
KW Hypothetical protein.
FT NON TER 143
SQ SEQUENCE 143 AA; 15689 MW; 1CD55C1EFA9B0AD8 CRC64;

Query Match 100.0%; Score 143; DB 10; Length 143;
Best Local Similarity 100.0%; Pred. No. 1.8e-148; Indels 0; Gaps 0;
Matches 143; Conservative 0; Mismatches 0;
QY 1 MVDNFPISQSTHAWCWSSTTRSPSRHHLRERIPCLALGVTAICSLVWIHVSHGGG 60
DB 1 MVDNFPISQSTHAWCWSSTTRSPSRHHLRERIPCLALGVTAICSLVWIHVSHGGG 60
QY 61 ISVALCSQCLOTNALRPPDCLTNGGCTGCGSLGHVDRFPQHSNEWNSGMDSCCKPLR 120
DB 61 ISVALCSQCLOTNALRPPDCLTNGGCTGCGSLGHVDRFPQHSNEWNSGMDSCCKPLR 120
QY 121 GFLUGVLTTPHPRKMEFAAIRAGKV 143

Query Match 5.6%; Score 8; DB 5; Length 271;
Best Local Similarity 100.0%; Pred. No. 3-2; 0; Gaps
Matches 8; Conservative 0; Mismatches 0; Indels 0;

DQ 21 SSTTRSPS 28
| | | | |
DB 233 SSTTRSPS 240

RESULT 3

ID Q9VUH3 PRELIMINARY; PRT; 295 AA.

AC Q9VUH3
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein Alr2373.
GN ALR2373.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2159285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
MA Tanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsuno M., Matsumoto A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
CYANOBACTERIUM ANABENA SP. STRAIN PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AF003589; BAB74072.1; --
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 295 AA; 31535 MW; 78769CADCBD6DCFCRC64;

Query Match 5.6%; Score 8; DB 16; Length 295;
Best Local Similarity 100.0%; Pred. No. 3-5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DQ 40 LALGVTAI 47
| | | | |
DB 178 LALGVTAI 185

RESULT 4

ID P93484 PRELIMINARY; PRT; 623 AA.

AC P93484
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-MAY-2003 (TrEMBLrel. 23, Last annotation update)
DE BP-80 vacuolar sorting receptor.
OS Pisum sativum (Garden pea)
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
OC eurosidia I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
OX NCBI_TaxID=3888;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Immature seed;
RX MEDLINE=94211872; PubMed=8159760;
RA Kirisch T., Paris N., Bucher J.M., Bevers L., Rogers J.C.;
RT Purification and initial characterization of a potential plant
RT vacuolar targeting receptor."
RL Proc. Natl. Acad. Sci. U.S.A. 91:3403-3407(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Immature seed;
RX MEDLINE=97451769; PubMed=9306690;
RA Paris N., Rogers S.W., Jiang L., Kirsch T., Bevers L., Phillips T.E.,
RA Rogers J.C.;

RT "Molecular cloning and further characterization of a probable plant
RL vacuolar sorting receptor.";
RL Plant Physiol. 115:29-39(1997).
DR EMBL; U79558; AAB72110.1; -
DR InterPro; IPR001152; Asx_Hydroxyl.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_Like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR003137; PA.
DR Pfam; PF02225; PA; 1.
DR SMART; SM00181; EGF; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS00840; PA; 1.
DR PROSITE; PS00840; PA; 1.
KW EGF-like domain; Receptor.
SQ SEQUENCE 623 AA; 68896 MW; 10CC08959BD08184 CRC64;

Query Match 5.6%; Score 8; DB 10; Length 623;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 CLTNNGGC 88
Db 415 CLTNNGGC 422

RESULT 5
C48662 PRELIMINARY; PRT; 624 AA.
AC C48662;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE PV72.
CS Cucurbita cv. Kurokawa Amakuri.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Cucurbitales; Cucurbitaceae; Cucurbita.
OC NCBI_TaxID=3666;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue-Developing cotyledon;
RX MEDLINE=98182843; PubMed=9522472;
RA Shamada T., Kuroyagagi M., Nishimura M., Kura-Nishimura I.;
RT "A pumpkin 72-kDa membrane protein of precursor-accumulating vesicles
RL Plant Cell Physiol. 38:1414-1420(1997).
DR EMBL; AB006809; BAA25079.1; -
DR InterPro; IPR001152; Asx_Hydroxyl.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_Like.
DR InterPro; IPR003137; PA.
DR Pfam; PF02225; PA; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS00840; PA; 1.
KW EGF-like domain.
SQ SEQUENCE 624 AA; 59028 MW; 71E81EDD24037FEP CRC64;

Query Match 5.6%; Score 8; DB 10; Length 624;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 CLTNNGGC 88
Db 417 CLTNNGGC 424

RESULT 6
Q8CFM6 PRELIMINARY; PRT; 1431 AA.

AC Q8CFM6;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hyaluronan receptor for endocytosis HARE precursor (Fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=22169209; PubMed=12181351;
RA Zhou B., Weigel J.A., Saxena A., Weigel P.H.;
RT "Molecular Cloning and Functional Expression of the Rat 175-kDa
RT Hyaluronan Receptor for Endocytosis.";
RL Mol. Biol. Cell 13:2853-2868(2002).
DR EMBL; AY007370; RAG13634.1; -
DR Receptor.
KW NON_TER
FT NON_TER
SQ SEQUENCE 1431 AA; 155914 MW; 3340A8F7AF366722 CRC64;

Query Match 5.6%; Score 8; DB 11; Length 1431;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 CLTNNGGC 88
Db 423 CLTNNGGC 430

RESULT 7
Q8R4U0 PRELIMINARY; PRT; 2559 AA.
ID Q8R4U0
AC Q8R4U0;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Stabilin-2.
GN STAB2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; Tissue=Liver;
RX MEDLINE=21818465; PubMed=11829752;
RA Politz O., Gratchev A., McCourt P.A.G., Schledzowski K., Guillot P.,
RA Johansson S., Svineng G., Franke P., Kannicht C., Kzyahkowska J.,
RA Longati P., Velten P.W., Johansson S., Goerd S.;
RT "Stabilin-1 and -2 constitute a novel family of fasciclin-like
RT hyaluronan receptor homologues.";
RL Biochem. J. 362:155-164(2002).
DR EMBL; AF364951; AAL91684.2; -
DR MGD; MGI:2178743; Stab2.
DR InterPro; IPR001589; Actbind actnin.
DR InterPro; IPR000782; BIGH3 fasciclin.
DR InterPro; IPR006209; EGF_Like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR000538; Link.
DR Pfam; PF00008; EGF; 18.
DR Pfam; PF02469; Fasciclin; 6.
DR Pfam; PF00193; Xlink; 1.
DR ProDom; PD000918; Link; 1.
DR SMART; SM00181; EGF; 21.
DR SMART; SM00554; FAS1; 7.
DR SMART; SM00445; LINK; 1.
DR PROSITE; PS00019; ACTININ 1; 1.
DR PROSITE; PS00022; EGF 1; 7.
DR PROSITE; PS01186; EGF_2; 16.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 2.

KW EGF-like domain.
SQ SEQUENCE 2559 AA; 277530 MW; 1C9855AD61BFF015 CRC64;

Query Match 5.6%; Score 8; DB 11; Length 2559;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 CLTNNGGC 88
DB 1566 CLTNNGGC 1573

RESULT 8
Q04390 PRELIMINARY; PRT; 64 AA.
AC Q04390; (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-NOV-1996 (TRENBLrel. 23, Last annotation update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Hypothetical protein (fragment).
OS Streptomyces lividans.
OC Bacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1916;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TK64;
RC Submitted (APR-1991) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-58 FROM N.A.
RC STRAIN=TK64;
RC MEDLINE=94010335; PubMed=7691688;
RA Sedlmeyer R., Linti G., Gregor K., Schmiegner H.,
RT "Sequences of tRNA-encoding genes and associated open reading frames
of Streptomyces lividans."
RL Gene 132:125-130(1993).
DR EMBL; X65556; CAA46527.1; --
FT NON_TPR 64
FT SEQUENCE 64 AA; 7142 MW; BC71B8310C1B1A26 CRC64;

Query Match 4.9%; Score 7; DB 2; Length 64;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 LALGVTA 46
DB 34 LALGVTA 40

RESULT 9
Q8XPHO PRELIMINARY; PRT; 124 AA.
AC Q8XPHO;
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Hypothetical signal peptide protein Rsp1670.
GN RSP1670 OR RS02229.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Ralstoniaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM1000;
RC MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave P., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brotier P., Camus J.C., Cattolico L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Siguler P., Thebaud P., Whalen M., Wincker P., Levy M.,

RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum."
RL Nature 415:497-502(2002).
DR EMBL; AL646086; CAD18821.1; --
KW Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 124 AA; 11776 MW; 9470F5AD3E781782 CRC64;

Query Match 4.9%; Score 7; DB 16; Length 124;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 LALGVTA 46
DB 16 LALGVTA 22

RESULT 10
Q9RCW2 PRELIMINARY; PRT; 174 AA.
AC Q9RCW2;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Putative membrane protein.
GN SCO0849 OR SCM2.02C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RC MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Harris D.E., Quail M.A., Kieser H.,
RA Thomson N.R., James K.D., Haxby S., Chandra G., Chen C.W., Collins M.,
RA Harper D., Bateman A., Brown S., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Cronin A., Fraser A., Goble A., Lark L., Murphy L., Oliver K., O'Neil S.,
RA Huang C.-H., Kieser T., Larke L., Murrells A., Rutherford K., Rutter S.,
RA Rabinowitz E., Rajandream M.A., Saunders R., Squares R., Taylor K.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warten T., Wietzorrek A., Woodward J., Barrall B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2)";
RL Nature 417:141-147(2002).
DR EMBL; AL939106; CAB65630.1; --
KW Complete proteome.
SQ SEQUENCE 174 AA; 18196 MW; FB8019102ACE6EBF CRC64;

Query Match 4.9%; Score 7; DB 16; Length 174;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 LALGVTA 46
DB 30 LALGVTA 36

RESULT 11
Q9XJD9 PRELIMINARY; PRT; 185 AA.
AC Q9XJD9;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE Hypothetical 21.5 kDa protein.
OS Streptococcus thermophilus bacteriophage DT1
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX NCBI_TaxID=90410;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DT1;
RC MEDLINE=99160757; PubMed=10049822;

```

Trenblay D.M., Moineau S.:
"complete genomic sequence of the lytic bacteriophage DT1 of
Streptococcus thermophilus.";
Virology 255:63-76(1999).
[2]
SEQUENCE FROM N.A.
STRAIN=DT1.
MEDLINE=21382762; PubMed=11489121;
Duplessis M., Moineau S.:
"Identification of a genetic determinant responsible for host
specificity in Streptococcus thermophilus bacteriophages.";
Mol. Microbiol. 41:325-336(2001).
[3]
SEQUENCE FROM N.A.
STRAIN=DT1.
Trenblay D.M., Moineau S.:
Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
EMBL; AF085222; AAD21922.1; -.
InterPro; IPR003615; HNH.nuc.
SMART; SM00507; HNHc; 1.
Hypothetical protein.
SEQUENCE 185 AA; 21532 MW; 68F8432CAFF0BE26 CRC64;

Query Match 4.9%; Score 7; DB 9; Length 185;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 CLALGVT 45
Db 99 CLALGVT 105
|||||

RESULT 12
Q8KMY5 PRELIMINARY; PRT; 204 AA.
ID Q8KMY5
AC Q8KMY5;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-NOV-2003 (TrEMBLrel. 23, Last annotation update)
DE "Hypothetical protein in frr 3' region (Fragment)."
GN YAEW
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OC NCBI_TaxID=562;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=K-12;
RC MEDLINE=82059454; PubMed=6272196;
RX An G., Bendtsen D.S., Mamelak L.A., Friesen J.D.;
RT "Organization and nucleotide sequence of a new ribosomal operon in
RT Escherichia coli containing the genes for ribosomal proteins S2 and
RT elongation factors 18.";
RL Nucleic Acids Res. 9:4163-4172(1981).
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN=K-12;
RC MEDLINE=83209630; PubMed=6343085;
RX Stephens P.E., Darlison M.G., Lewis H.M., Guest J.R.;
RT "The pyruvate dehydrogenase complex of Escherichia coli K12.
RT Nucleotide sequence encoding the pyruvate dehydrogenase component.";
RL Eur. J. Biochem. 133:155-162(1983).
RN [3]
RN SEQUENCE FROM N.A.
RP STRAIN=K-12;
RC MEDLINE=63234434; PubMed=6345153;
RX Stephens P.E., Darlison M.G., Lewis H.M., Guest J.R.;
RT "The pyruvate dehydrogenase complex of Escherichia coli K12.
RT Nucleotide sequence encoding the dihydrolipoamide acetyltransferase
RT component.";
RL Eur. J. Biochem. 133:481-489(1983).
RN [4]
RN SEQUENCE FROM N.A.
RP

```

RP
RC-
RC STRAIN-K-12;
MEDLINE=88058963; PubMed=3316212;
Tabor C.W., Tabor H.;
"The speBspeD Operon of Escherichia coli: Formation and processing of
a proenzyme form of S-adenosylmethionine decarboxylase.";
J. Biol. Chem. 262:16037-16040(1987).
[14]
RN
RP SEQUENCE FROM N.A.
RN
RP STRAIN-K-12;
MEDLINE=84152237; PubMed=2450046;
Gebhard W., Schreitmüller T., Hochstrasser K.;
"Complementary DNA and derived amino acid sequence of the precursor of
one of the three protein components of the inter-alpha-trypsin
inhibitor complex.";
FEBS Lett. 229:63-67(1988).
[15]
RN
RP SEQUENCE FROM N.A.
RN
RP STRAIN-K-12;
MEDLINE=89061679; PubMed=2904262;
Andrews S.C., Guest J.R.;
"Nucleotide sequence of the gene encoding the GMP reductase of
Escherichia coli K12";
Biochem. J. 255:35-43(1988).
[16]
RN
RP SEQUENCE FROM N.A.
RN
RP STRAIN-K-12;
MEDLINE=88227880; PubMed=3372485;
Mellano M.A., Cooksey D.A.;
"Nucleotide sequence and organization of copper resistance genes from
Pseudomonas syringae pv. tomato.";
J. Bacteriol. 170:2879-2883(1988).
[17]
RN
RP SEQUENCE FROM N.A.
RN
RP STRAIN-K-12;
MEDLINE=89155419; PubMed=2537812;
Liu J., Parkinson J.S.;
"Genetics and sequence analysis of the pcnB locus, an Escherichia coli
gene involved in plasmid copy number control.";
J. Bacteriol. 171:1254-1261(1988).
[18]
RN
RP SEQUENCE FROM N.A.
RN
RP STRAIN-K-12;
MEDLINE=89008347; PubMed=3049588;
Sung Y., Fuchs J.A.;
"Characterization of the cyn operon in Escherichia coli K12.";
J. Biol. Chem. 263:14769-14775(1988).
[19]
RN
RP SEQUENCE FROM N.A.
RN
RP STRAIN-K-12;
MEDLINE=89057448; PubMed=3057437;
Lipinska B., Sharma S., Georgopoulos C.;
"Sequence analysis and regulation of the hcrA gene of Escherichia
coli: A sigma-32-independent mechanism of heat-inducible
transcription.";
Nucleic Acids Res. 16:10053-10067(1988).
[20]
RN
RP SEQUENCE FROM N.A.
RN
RP STRAIN-K-12;
MEDLINE=90128278; PubMed=2693214;
Roncero M.I., Jepseen L.P., Stroman P., van Heeswijk R.;
"Characterization of a leuA gene and an ARS element from Mucor
circinelloides.";
Gene 84:335-343(1989).
[21]
RN
RP SEQUENCE FROM N.A.
RN
RP STRAIN-K-12;
MEDLINE=89227155; PubMed=2666401;
Xie Q.W., Tabor C.W., Tabor H.;
"Spermidine biosynthesis in Escherichia coli the promoter and the
termination regions of the speed operon.";
J. Bacteriol. 171:4457-4465(1989).
[22]

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021.
RX MEDLINE=21336508; PubMed=11481431;
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
RA Vorhoefer F.J., Hernandez-Lucas I., Becker A., Cowie J.,
RA Golding B., Puehler A.;
RT "The complete sequence of the 1.683-kb pSymb megaplasmid from the N2-
fixing endosymbiont Sinorhizobium meliloti.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
DR EMBL: AL603644; CAC49058.1;
DR InterPro: IPR000595; cNMP_binding.
DR Pfam: PF00027; cNMP_binding; 1.
DR SMART: SM00100; cNMP; 1.
KW Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 243 AA; 26741 MW; 698B4EFC65285162 CRC64;

Query Match 4.9%; Score 7; DB 16; Length 243;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 41 ALGVTAI 47
Db 190 ALGVTAI 196
|||||
|

RESULT 15
Q888Q9 PRELIMINARY; PRT; 244 AA.
AC Q888Q9, 2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein.
OS Tobacco bushy top virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Umbravirus.
OX NCBI_TaxID=184020;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Baoshan;
RA Mo X., Qin X., Tan Z., Li T., Wu J., Chen H.;
RT "First Report of Tobacco Bushy Top Disease in China.";
RL Plant Dis. 86:74-74(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Baoshan;
RA Mo X., Qin X., Li T., Wu J., Chen H.;
RT "Complete sequence and genome organization of Tobacco bushy top
virus.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF402620; AAN62864.1;
KW Hypothetical protein.
SQ SEQUENCE 244 AA; 26895 MW; 1EE8BC4223EE96D5 CRC64;

Query Match 4.9%; Score 7; DB 12; Length 244;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 124 LGVLTTPH 130
Db 32 LGVLTTPH 38
|||||
|

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 13:22:10 ; Search time 17 Seconds
(without alignment)

395.578 Million cell updates/sec

Title: US-09-868-025-2

Perfect score: 143

Sequence: 1 MVDTFPISEQSETHAWCMS.....LGVLPHPXMEFAAIRAGKV 143

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	4.9	139	Y241 ARCFU	O29998 archaeoglob
2	7	4.9	205	BKUA CHLMU	Q9p109 chlamydia m
3	7	4.9	355	BUK LISIN	Q92by5 listeria in
4	7	4.9	355	BUK LISNO	O8v7b6 listeria m
5	7	4.9	359	CD73 HUMAN	P21654 hunc sapien
6	7	4.9	398	DXR ECOS7	O8x8v1 escherichia
7	7	4.9	398	DXR ECOLI	P45568 escherichia
8	7	4.9	398	DXR SALT1	O829a6 salmonella
9	7	4.9	398	DXR SALT2	O829p3 salmonella
10	7	4.9	398	DXR VERPE	O8zh62 versinia pe
11	7	4.9	659	SVT THERH	P56881 thermus the
12	7	4.9	982	RRPO GGNV	Q993m1 greasy grou
13	6	4.2	83	HEPC MOUSE	Q99mh3 rattus norv
14	6	4.2	84	HEPC RAT	P10817 rattus norv
15	6	4.2	94	COXD RAT	P29902 paracoccus
16	6	4.2	105	YMOX FARDE	Q20779 caenorhabdi
17	6	4.2	128	COXE CAEEL	P34401 caenorhabdi
18	6	4.2	131	YLU7 CAEEL	P31734 xanthomonas
19	6	4.2	143	GSFG XANCP	O9t1t5 bacterioph
20	6	4.2	146	LYCV BPAPS	O8dmo1 synchococc
21	6	4.2	152	SSRP SYNEL	O2586 streptomyce
22	6	4.2	154	CIRC STRCN	P74355 synchocyst
23	6	4.2	154	SSRP SYNY3	O8ym70 anabaena sp
24	6	4.2	155	SRP SYNY3	O8ra24 thermosae
25	6	4.2	184	RRP THERN	O03892 notropocra
26	6	4.2	198	COX2 NOTPE	O03895 tinamus maj
27	6	4.2	198	COX2 TINNA	O03889 apteryx aus
28	6	4.2	199	COX2 APTAU	O03890 casuarus b
29	6	4.2	199	COX2 CASPE	O03891 dronaeus no
30	6	4.2	199	COX2 DRONO	O03893 rhea ameri
31	6	4.2	199	COX2 RHEAM	P77736 escherichia
32	6	4.2	201	YAHB ECOLI	O84033 chlamydia t
33	6	4.2	205	KGUA CHLTR	

34 6 4.2 208 1 NARP HAEIN
35 6 4.2 208 1 NORD NEIMA
36 6 4.2 208 1 NORD NEIME
37 6 4.2 209 1 CDBB SAISC
38 6 4.2 216 1 MSRA XANAC
39 6 4.2 216 1 MSRA XANAC
40 6 4.2 216 1 MSRA XANCP
41 6 4.2 223 1 PYRE RHILU
42 6 4.2 227 1 COX2 CHICK
43 6 4.2 227 1 COX2 COTJA
44 6 4.2 228 1 COX2 ANAPL
45 6 4.2 228 1 COX2 CAIMO

ALIGNMENTS

RESULT 1
ID Y241 ARCFU STANDARD; PRT; 139 AA.
AC O29998;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AF0241.
GN AF0241.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
Richardson D.L., Kerlavage A.R., Graham D.E., Kyrtides N.C.,
Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Uterback T.,
Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
Sadov P.M., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
Venter J.C.;
*The complete genome sequence of the hyperthermophilic, sulphate-
reducing archaeon Archaeoglobus fulgidus*;
RL Nature 390:364-370(1997)
CC - SIMILARITY: BELONGS TO THE UPF0066 (VIRR) FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC ENBL; AB001089; AAB90992.1; -
DR PIR; A69280; A69280.
DR TIGR; AF0241; -
DR InterPro; IPR001378; UPF0066.
DR Pfam; PF01980; UPF0066; 1.
DR ProDom; PD006705; UPF0066; 1.
DR TIGRPFams; TIGR00104; TIGR00104; 1.
DR PROSITE; PS01318; UPF0066; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 139 AA; 15809 MW; A4E98224A1FA841 CRC64;

Query Match 4.9%; Score 7; DB 1; Length 139;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 TTRSPR 29
DB 80 TTRSPR 86

RESULT 2
KGUA_CHLMO STANDARD; PRT; 205 AA.
ID Q9PL09; DT 16-OCT-2001 (Rel. 40, Created)
AC 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Guanylate_kinase (EC 2.7.4.8) (GMP kinase).
GN GWK OR TC0299.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiales; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MoPn / Nigg;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
CC 1- FUNCTION: ESSENTIAL FOR RECYCLING GMP AND INDIRECTLY, GOMP.
CC 1- CATALYTIC ACTIVITY: ATP + GMP = ADP + GDP.
CC 1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC 1- SIMILARITY: BELONGS TO THE GUANYLATE KINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF002297; AAF39164.1;
CC PIR; F81718; F81718.
CC HSP; P15454; 1GKY.
CC TIGR; TC0299; 1.
CC HAMAP; MF_00328; 1.
CC InterPro; IPR000619; Guanylate_kin.
CC Pfam; PF00625; Guanylate_kin; 1.
CC SMART; SM00072; GUKC; 1.
CC PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
CC PROSITE; PS00052; GUANYLATE_KINASE_2; 1.
CC TRANSFERASE; Kinase; ATP-binding; Complete proteome.
KW NP_BIND; 25 32 ATP; (BY SIMILARITY).
FT SEQUENCE 205 AA; 22977 MW; E0CCD5DDA227BB3 CRC64;
Query Match 4.9%; Score 7; DB 1; Length 205;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 21 SSTTRSP 27
DB 51 SSTTRSP 57

RESULT 3
BUX_LISIN STANDARD; PRT; 355 AA.
ID BUX_LISIN
AC Q92BY5; DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable butyrate kinase (EC 2.7.2.7) (BK) (Branched-chain carboxylic

acid kinase).
DE BUK OR LIN1407.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / Serovar 6a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero P., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fathi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunat F., Kurapkut G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
CC 1- CATALYTIC ACTIVITY: ATP + 2-butanolate = ADP + butanoyl phosphate.
CC 1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC 1- SIMILARITY: Belongs to the acetokinase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL596168; CAC96638.1;
CC PIR; APL608; APL608.
CC List1; LIN01407; 1.
CC HAMAP; MF_00542; 1.
CC InterPro; IPR000890; Acetate_kin.
CC Pfam; PF00871; Acetate_kinase; 1.
CC PRINTS; PR00471; ACETATE_KINASE.
CC PROSITE; PS01075; ACETATE_KINASE_1; 1.
CC PROSITE; PS01076; ACETATE_KINASE_2; 1.
CC TRANSFERASE; Kinase; Complete proteome.
KW SEQUENCE 355 AA; 38891 MW; B1B5C3DE98169E5 CRC64;
Query Match 4.9%; Score 7; DB 1; Length 355;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 59 GGGISVA 64
DB 185 GGGISVA 191

RESULT 4
BUX_LISMO STANDARD; PRT; 355 AA.
ID BUX_LISMO
AC Q8Y7B6; DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable butyrate kinase (EC 2.7.2.7) (BK) (Branched-chain carboxylic
acid kinase).
GN BUK OR LMO1370.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD-e / Serovar 1/2a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,

BA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
 RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
 RA Domann E., Dominguez-Bernal G., Duchaud A., Durant L., Dussurget O.,
 RA Entian K.-D., Feihl H., Garcia-del Portillo F., Garrido P.,
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
 RA Jones L.-M., Kaerst U., Krefz J., Kuhn M., Kunst F., Kurapat G.,
 RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
 RA Nordstiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
 RA Remmel B., Rose M., Schluter T., Smoes N., Tierrez A.,
 RA Vazquez-Boland J.-A., Voss H., Wehlend J., Cossart P.,
 RA "Comparative genomics of *Listeria species*,"
 RT Science 294:949-952(2001).
 CC -1- CATALYTIC ACTIVITY: ATP + 2-butanolate = ADP + butanoyl phosphate.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: Belongs to the acetokinase family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AL591978; CAC99448.1; -;
 CC PIR; AB1246; AB1246.
 CC Listlist; LMO01370; -;
 CC HANAP; MF_00542; -; 1.
 DR InterPro; IPR000890; Acetate_kin.
 DR Pfam; PF00871; Acetate_kinase; 1.
 DR PRINTS; PR00471; ACETATEKINASE.
 DR PROSITE; PS01075; ACETATE KINASE 1; 1.
 DR PROSITE; PS01076; ACETATE KINASE 2; 1.
 KW Transferase; Kinase; Complete proteome.
 SQ SEQUENCE 355 AA; 38969 MW; DABBPFC22118E8A CRC64;
 Query Match 4.9%; Score 7; DB 1; Length 355;
 Best Local Similarity 100.0%; Pred. No. 9.7;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 58 GGGISVA 64
 DB 185 GGGISVA 191
 |||||
 RESULT 5
 CD72 HUMAN
 ID CD72 HUMAN STANDARD; PRT; 359 AA.
 AC P21854;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 15-FEB-2003 (Rel. 42, Last annotation update)
 DE B-cell differentiation antigen CD72 (Lyb-2).
 GN CD72.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. PubMed=2141045;
 RX MEDLINE=90278102; PubMed=12477932;
 RA von Hoegen I., Nakayama E., Parnes J.R.;
 RT "Identification of a human protein homologous to the mouse Lyb-2 B
 cell differentiation antigen and sequence of the corresponding
 RT cDNA."
 RL J. Immunol. 144:4870-4877(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancreas, and Spleen;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.P., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
 RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Murley D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinaki M.I., Skaleka U., Smalhus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences,"
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -1- FUNCTION: PLAYS A ROLE IN B CELL PROLIFERATION AND
 CC DIFFERENTIATION. ASSOCIATES WITH CD5.
 CC -1- SUBUNIT: Homodimer; disulfide-linked.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein.
 CC -1- TISSUE SPECIFICITY: PRE-B-CELLS AND B-CELLS BUT NOT TERMINALLY
 CC DIFFERENTIATED PLASMA CELLS.
 CC -1- SIMILARITY: Contains 1 C-type lectin family domain.
 CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD72 entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd72.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M54992; AA36189.1; -;
 CC EMBL; BC030227; AA360227.1; -;
 CC PIR; A43532; A43532.
 CC Genew; HGNC:1696; CD72.
 CC MIM; 107272; -;
 DR GO; GO:0005102; P:receptor binding activity; TAS.
 DR GO; GO:0007155; P:cell adhesion; TAS.
 DR InterPro; IPR001304; Lectin_C.
 DR Pfam; PF00059; lectin_C; 1.
 DR SMART; SM00034; CLECT; 1.
 DR PROSITE; PS00615; C-TYPE LECTIN 1; FALSE_NEG.
 DR PROSITE; PS0041; C-TYPE LECTIN 2; 1.
 KW Antigen; B-cell; Glycoprotein; Transmembrane; Lectin; Signal-anchor.
 FT DOMAIN 1 95 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 96 116 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 117 359 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 232 352 C-TYPE LECTIN (LONG FORM).
 FT DISULFID 233 244 BY SIMILARITY.
 FT DISULFID 261 350 BY SIMILARITY.
 FT DISULFID 325 342 BY SIMILARITY.
 FT CARBOHYD 136 136 N-LINKED (GLCNAC... (POTENTIAL).
 SQ SEQUENCE 359 AA; 40220 MW; 87A52C028AC17E44 CRC64;
 Query Match 4.9%; Score 7; DB 1; Length 359;
 Best Local Similarity 100.0%; Pred. No. 9.8;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 42 LGVTAIC 48
 DB 107 LGVTAIC 113
 |||||
 RESULT 6
 DXR_ECO57
 ID DXR_ECO57 STANDARD; PRT; 398 AA.
 AC Q8X8V1;
 DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE 1-deoxy-D-xylulose 5-phosphate reductoisomerase (EC 1.1.1.267) (DXP
 DE reductoisomerase) [1-deoxyxylulose-5-phosphate reductoisomerase].
 GN DXR OR Z0184 OR EC50125
 GN Escherichia coli O157:H7.
 OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potancus K.,
 RA Apodaca J., Ananthan N.T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
 RT Nature 409:529-533(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 050952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sakakawa C., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 RT O157:H7 and genomic comparison with a laboratory strain K-12";
 RL DNA Res. 8:11-22(2001).
 CC -!- FUNCTION: Catalyzes the NADP-dependent rearrangement and reduction
 CC of 1-deoxy-D-xylulose 5-phosphate (DXP) to 2-C-methyl-D-erythritol
 CC 4-phosphate (MEP) (by similarity).
 CC -!- CATALYTIC ACTIVITY: 2-C-methyl-D-erythritol 4-phosphate + NADP(+) +
 CC = 1-deoxy-D-xylulose 5-phosphate + NADPH.
 CC -!- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; second
 CC step.
 CC -!- SIMILARITY: BELONGS TO THE DXR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>
 CC or send an email to license@ebi.ac.uk).
 CC -----
 DR EMBL; AF005193; AAG54475.1; -;
 DR EMBL; AP002550; BAB33598.1; -;
 DR FIR; G85501; G85501.
 DR FIR; G90650; G90650.
 DR HAMAP; MF_00183; -; 1.
 DR InterPro; IPR003821; DXP reductoisomerase.
 DR Pfam; PF02670; DXP reductoisom; 1.
 DR TIGRFAMs; TIGR00243; Dxr; 1.
 KW Isoprene biosynthesis; Oxidoreductase; NADP; Complete proteome.
 FT NP_BIND 7 14 NADPH (POTENTIAL).
 SQ SEQUENCE 399 AA; 43361 MW; 88532683A4FF082E CRC64;
 Query Match 4.9%; Score 7; DB 1; Length 398;
 Best Local Similarity 100.0%; Pred.No. 11;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 136 RAIRAGK 142
 DB 112 AAIRAGK 118
 RESULT 7
 DXR_ECOLI
 ID DXR_ECOLI STANDARD; PRT; 398 AA.

AC P45568; P77209;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE 1-deoxy-D-xylulose 5-phosphate reductoisomerase (EC 1.1.1.267) (DXP
 DE reductoisomerase) [1-deoxyxylulose-5-phosphate reductoisomerase].
 GN DXR OR B0173.
 GN Escherichia coli.
 OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A. AND CHARACTERIZATION.
 RC STRAIN=K12 / W3110;
 RX MEDLINE=98174274; PubMed=9707569;
 RA Takahashi S., Kuzuyama T., Watanabe H., Seto H.;
 RT "A 1-deoxy-D-xylulose 5-phosphate reductoisomerase catalyzing the
 RT formation of 2-C-methyl-D-erythritol 4-phosphate in an alternative
 RT nonmevalonate pathway for terpenoid biosynthesis";
 RT Proc. Natl. Acad. Sci. U.S.A. 95:9879-9884(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12";
 RT Science 277:1453-1474(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / W3110;
 RX Takenoto K., Mori H., Murayama N., Kataoka K., Yano M., Itoh T.,
 RA Yamanoto Y., Inokuchi H., Miki T., Hatada E., Fukuda R., Ichihara S.,
 RA Mizuno T., Makino K., Nakata A., Yura T., Sampei G., Mizobuchi K.;
 RT "Systematic sequencing of the Escherichia coli genome: analysis of the
 RT 4.0 - 6.0 min (189,987 - 281,416bp) region";
 RT Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,
 RA Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
 RA Lashari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,
 RA Davis R.W.;
 RT Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 1-49 FROM N.A.
 RX MEDLINE=93077430; PubMed=1447125;
 RA Yamanaka K., Ogura T., Niki H., Hiraga S.;
 RT "Identification and characterization of the smba gene, a suppressor
 RT of the mukB null mutant of Escherichia coli";
 RL J. Bacteriol. 174:7517-7526(1992).
 RN [6]
 RP IDENTIFICATION.
 RX MEDLINE=96032851; PubMed=7567469;
 RA Borodovsky M., McIninch J., Koonin E.V., Rudd K.E., Medigue C.,
 RA Danchin A.;
 RT "Detection of new genes in a bacterial genome using Markov models for
 RT three gene classes";
 RL Nucleic Acids Res. 23:3554-3562(1995).
 RN [7]
 RP CHARACTERIZATION.
 RX MEDLINE=20123893; PubMed=10631325;
 RA Radkiewicz T., Rohdich P., Wungskintaweekul J., Herz S., Kis K.,
 RA Eisenreich W., Bacher A., Zenk M.H., Arigoni D.;
 RT "Biosynthesis of terpenoids: 1-deoxy-D-xylulose-5-phosphate
 RT reductoisomerase from Escherichia coli is a class B dehydrogenase";
 RL FEBS Lett. 465:157-160(2000).
 RN [8]
 RP CHARACTERIZATION, AND MUTAGENESIS.
 RX MEDLINE=20347905; PubMed=10787409;
 RA Kuzuyama T., Takahashi S., Takagi M., Seto H.;

RL Nature 413:523-527(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KIM5 / Biovar Mediaevalis;
 RA MEDLINE=22137863; PubMed=12142430;
 RX Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
 RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
 RA Fetherston J.D., Linder L.E., Brubaker R.R., Plano G.V.,
 RA Straley S.C., McDonough K.A., Nilles M.L., Watson J.S., Blattner F.R.,
 RA Perry R.D.;
 RT "Genome sequence of *Yersinia pestis* KIM.";
 RL J. Bacteriol. 184:4601-4611(2002).
 CC -!- FUNCTION: Catalyzes the NADP-dependent rearrangement and reduction
 CC of 1-deoxy-D-xylulose-5-phosphate (DXP) to 2-C-methyl-D-erythritol
 CC 4-phosphate (MEP) (By similarity).
 CC -!- CATALYTIC ACTIVITY: 2-C-methyl-D-erythritol 4-phosphate + NADP(+) = 1-deoxy-D-xylulose 5-phosphate + NADPH.
 CC -!- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; second step.
 CC -!- SIMILARITY: BELONGS TO THE DXR FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AJ141414; CAC989890.1;
 CC EMBL; AE013914; AAM86681.1;
 CC PIR; AG0128; AG0128.
 CC HAMAP; MF_00183; -; 1.
 CC InterPro; IPR003821; DXP_reductoisomerase.
 CC Pfam; PF02670; DXP_reductoisom; 1.
 CC TIGRfam; TIGR00243; Dxr; 1.
 CC Isoprene biosynthesis; Oxidoreductase; NADP; Complete proteome.
 CC NP_BIND 7 14 NADPH (POTENTIAL).
 CC SEQUENCE 398 AA; 43115 MW; C9B1FC9E0166D057 CRC64;
 CC
 CC Query Match 4.9%; Score 7; DB 1; Length 398;
 CC Best Local Similarity 100.0%; Pred.No.11;
 CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 136 AIRAGK 142
 CC Db 112 AIRAGK 118
 CC
 CC RESULT 11
 CC SYT THETH STANDARD; PRT; 659 AA.
 CC AC P5681;
 CC DT 30-MAY-2000 (Rel. 39, Created)
 CC DT 30-MAY-2000 (Rel. 39, Last sequence update)
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
 CC DE Threonyl-tRNA synthetase (EC 6.1.1.3) (Threonine--tRNA ligase)
 CC (Thrs).
 CC GN THRS.
 CC OS Thermus thermophilus.
 CC OC Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
 CC OC Thermus.
 CC OX NCBI_TaxID=274;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RX STRAIN=HB8 / ATCC 27634;
 CC MEDLINE=20098514; PubMed=10632708;
 CC Cura V., Moras D., Kern D.;
 RA "Sequence analysis and modular organization of threonyl-tRNA
 RT synthetase from *Thermus thermophilus* and its interrelation with
 RT threonyl-tRNA synthetases of other origins.";
 RL Eur. J. Biochem. 267:379-393(2000).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Florea D., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium
 RT LT2.";
 RL Nature 413:852-856(2001).
 CC -!- FUNCTION: Catalyzes the NADP-dependent rearrangement and reduction
 CC of 1-deoxy-D-xylulose-5-phosphate (DXP) to 2-C-methyl-D-erythritol
 CC 4-phosphate (MEP) (By similarity).
 CC -!- CATALYTIC ACTIVITY: 2-C-methyl-D-erythritol 4-phosphate + NADP(+) = 1-deoxy-D-xylulose 5-phosphate + NADPH.
 CC -!- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; second step.
 CC -!- SIMILARITY: BELONGS TO THE DXR FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AE008705; AAL19184.1;
 CC StyGene; SG7???; dxr.
 CC HAMAP; MF_00183; -; 1.
 CC InterPro; IPR003821; DXP_reductoisomerase.
 CC Pfam; PF02670; DXP_reductoisom; 1.
 CC TIGRfam; TIGR00243; Dxr; 1.
 CC Isoprene biosynthesis; Oxidoreductase; NADP; Complete proteome.
 CC NP_BIND 7 14 NADPH (POTENTIAL).
 CC SEQUENCE 398 AA; 43352 MW; CB720D959C3F8FCC CRC64;
 CC
 CC Query Match 4.9%; Score 7; DB 1; Length 398;
 CC Best Local Similarity 100.0%; Pred.No.11;
 CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 136 AIRAGK 142
 CC Db 112 AIRAGK 118
 CC
 CC RESULT 10
 CC DXR_YERPE STANDARD; PRT; 398 AA.
 CC AC Q8ZH62; PRT; 398 AA.
 CC DT 28-FEB-2003 (Rel. 41, Created)
 CC DT 28-FEB-2003 (Rel. 41, Last sequence update)
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
 CC DE 1-deoxy-D-xylulose 5-phosphate reductoisomerase (EC 1.1.1.267) (DXP
 CC reductoisomerase) (1-deoxyxylulose-5-phosphate reductoisomerase).
 CC DXR OR YP01048 OR Y1313.
 CC GN *Yersinia pestis*.
 CC OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC OC Enterobacteriaceae; *Yersinia*.
 CC OX NCBI_TaxID=632;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RX STRAIN=CO-92 / Biovar Orientalis;
 CC MEDLINE=21470413; PubMed=11586360;
 CC Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
 RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
 RA Baker S., Baugham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Feltwell T., Hamlin N., Hollroyd S., Jagels K., Karlyshev A.V.,
 RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
 RT "Genome sequence of *Yersinia pestis*, the causative agent of plague.";

RP SEQUENCE OF 1-45, AND CHARACTERIZATION.
RC STRAIN=HB8 / ATCC 27634;
RX MEDLINE=94304998; PubMed=8031907;
RA Zheltosova J., Melnikova E., Garber M., Reinbolt J., Kern D.,
RT "Ehresmann C., Ehresmann B.,
RA "Phreonyl-tRNA synthetase from *Thermus thermophilus*: purification and
RT some structural and kinetic properties.";
RL Biochimie 76:71-77(1994)
CC -!- FUNCTION: Replicates the viral genome which is composed of two RNA
CC segments, RNA1 and RNA2.
CC -!- CATALYTIC ACTIVITY: ATP + L-threonine + tRNA(Thr) = AMP +
CC diphosphate + L-threonyl-tRNA(Thr).
CC -!- SUBUNIT: Homodimer.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- DOMAIN: THE C-TERMINAL DOMAIN RECOGNIZES THE ANTICODON BASES.
CC -!- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AJ250953; CAB65493.1; -
CC DR HSSP: P00955; 1EVL.
CC DR HAMAP: MF 00184; -; 1.
CC DR InterPro: IPR004154; HGTP anticodon.
CC DR InterPro: IPR004095; TGS Dom.
CC DR InterPro: IPR002314; tRNA-synt 2b.
CC DR InterPro: IPR002320; tRNA-synt_thr.
CC DR InterPro: IPR006195; tRNA_ligase_1.
CC Pfam: PF03129; HGTP anticodon; 1.
CC Pfam: PF02824; TGS; 1.
CC Pfam: PF00587; tRNA-synt 2b; 1.
CC PRINTS: PR01047; TRNASYNTHTHR.
CC TIGRfam: TIGR00418; ttrs; 1.
CC PROSITE: PS50862; AA tRNA_LIGASE II; 1.
CC AMINOACYL-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
CC Metal-binding; Zinc.
FT DOMAIN 234 548 CATALYTIC.
FT DOMAIN 338 341 POLY-GLU.
FT METAL 349 349 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 400 400 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 529 529 ZINC (CATALYTIC) (BY SIMILARITY).
FT CONFLICT 42 42 E - T (IN REF. 2). CRC64;
SQ SEQUENCE 659 AA; 75337 MW; 1993ALBF115FF4 CRC64;
Query Match 4.9%; Score 7; DB 1; Length 659;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 121 GEFGLVL 127
DB 410 GEFGLVL 416
RESULT 12
RRPO GGNV
ID RRPO GGNV STANDARD; PRT; 982 AA.
AC Q933M1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE RNA-directed RNA polymerase (EC 2.7.7.48) (Rdrp) (RNA replicase)
DE (Protein A).
OS Greasy grouper nervous necrosis virus (GGNV) (Epinephelus tauvina
OS nervous necrosis virus).
OC Viruses, ssRNA positive-strand viruses, no DNA stage; Nodaviridae;
OC Betanodavirus.
OX NCBI_TaxID=143921;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=Singapore;
RX MEDLINE=21103005; PubMed=11172107;
RA Tan C., Huang B., Chang S.F., Ngoh G.H., Munday B., Chen S.C.,
RA Kwang J.;
RT "Determination of the complete nucleotide sequences of RNA1 and RNA2
RT from greasy grouper (*Epinephelus tauvina*) nervous necrosis virus,
RT Singapore strain.";
RL J. Gen. Virol. 82:647-653(2001).
CC -!- FUNCTION: Replicates the viral genome which is composed of two RNA
CC segments, RNA1 and RNA2.
CC -!- CATALYTIC ACTIVITY: Nucleoside triphosphate = N diphosphate +
CC (RNA) (N).
CC -!- SIMILARITY: BELONGS TO THE NODAVIRUSES RNA POLYMERASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF319555; AAK21877.1; -
CC DR InterPro: IPR007095; RNA_pol_DS_PS.
CC DR InterPro: IPR007094; RNA_pol_PSVir.
CC KW Transferase; RNA-directed RNA polymerase.
CC SQ SEQUENCE 982 AA; 110395 MW; CC61FB47B75CE0DA CRC64;
Query Match 4.9%; Score 7; DB 1; Length 982;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 43 GVTAICS 49
DB 231 GVTAICS 237
RESULT 13
HEPC_MOUSE
ID HEPC_MOUSE STANDARD; PRT; 83 AA.
AC Q9EQ21;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hecpudin precursor.
DE HAMP OR HEPC.
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND SUBCELLULAR LOCATION.
RX MEDLINE=21269329; PubMed=11113132;
RA Pigeon C., Ilyin G., Courseaud B., Leroyer P., Turlin B., Brissot P.,
RA Loreal O.;
RA "A new mouse liver-specific gene, encoding a protein homologous to
RA human antimicrobial peptide hepcidin, is overexpressed during iron
RA overload.";
RL J. Biol. Chem. 276:7811-7819(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Sapich M., Soares M.B., Donaldson M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalon D.K., Munzy D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahney J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 RL human and mouse cDNA sequences.",
 RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RP POSSIBLE FUNCTION IN IRON HOMEOSTASIS.
 RX MEDLINE=21353006; PubMed=11447267;
 RA Nicolas G., Bennoun M., Devaux I., Beaumont C., Grandchamp B.,
 RA Kahn A., Vaulont S.,
 RT "Lack of hepcidin gene expression and severe tissue iron overload in
 RT upstream stimulatory factor 2 (USF) knockout mice.",
 RL Proc. Natl. Acad. Sci. U.S.A. 98:8780-8785(2001).
 CC -1- FUNCTION: Seems to act as a signaling molecule involved in the
 CC maintenance of iron homeostasis. Seems to be required in
 CC conjunction with HFE to regulate both intestinal iron absorption
 CC and iron storage in macrophages. May also have antimicrobial
 CC activity.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: Expressed in liver.
 CC -1- SIMILARITY: BELONGS TO THE HEPICIDIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch.)
 CC
 CC EMBL; AF297664; AAG49293.1; -
 CC DR EMBL; BC021587; AAH21587.1; -
 CC DR MGD; MGI:1933533; Hamp.
 CC DR GO; GO:0003797; P:antibacterial peptide activity; IDA.
 CC DR GO; GO:0003799; F:antifungal peptide activity; IDA.
 CC
 CC SIGNAL 1 23 POTENTIAL.
 CC ANTIBIOTIC; Hormone; Signal.
 CC PROPEP 24 53
 CC CHAIN 59 83
 CC DISULFID 65 71
 CC DISULFID 68 81
 CC DISULFID 69 80
 CC DISULFID 72 77
 CC DISULFID 72 77
 CC SEQUENCE 83 AA; 9352 MW; C76423EA46260B18 CRC64;
 CC
 CC Query Match 4.2%; Score 6; DB 1; Length 83;
 CC Best Local Similarity 100.0%; Pred. No. 33;
 CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 3 DTNPFI 8
 CC DB 59 DTNPFI 64
 CC
 CC RESULT 14
 CC HEPC_RAT
 CC ID HEPC_RAT STANDARD; PRT; 84 AA.
 CC AC Q99MH3;
 CC DT 28-FEB-2003 (Rel. 41, Created)
 CC DT 28-FEB-2003 (Rel. 41, Last sequence update)
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
 CC DE Hepcidin precursor.
 CC DE Hamp OR HEPC.
 CC GN HAMP OR HEPC.
 CC OS Rattus norvegicus (Rat).
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 CC OC NCBI_TaxID=10116;
 CC RN [1]

RP SEQUENCE FROM N.A.
 RX STRAIN=Sprague-Dawley; TISSUE=Liver;
 CC MEDLINE=21269329; PubMed=1113132;
 RA Pigeon C., Ilyin G., Courseaud B., Leroyer P., Turlin B., Briesot P.,
 RA Loreal O.,
 RT "A new mouse liver-specific gene, encoding a protein homologous to
 RT human antimicrobial peptide hepcidin, is overexpressed during iron
 RT overload.",
 RL J. Biol. Chem. 276:7811-7819(2001).
 CC -1- FUNCTION: Seems to act as a signaling molecule involved in the
 CC maintenance of iron homeostasis. Seems to be required in
 CC conjunction with HFE to regulate both intestinal iron absorption
 CC and iron storage in macrophages. May also have antimicrobial
 CC activity (By similarity).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO THE HEPICIDIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch.)
 CC
 CC EMBL; AF344185; AAK12966.1; -
 CC DR EMBL; AF344185; AAK12966.1; -
 CC KW Antibiotic; Hormone; Signal.
 CC SIGNAL 1 23 POTENTIAL.
 CC PROPEP 24 54
 CC CHAIN 60 84
 CC DISULFID 66 72
 CC DISULFID 69 82
 CC DISULFID 70 81
 CC DISULFID 73 78
 CC SEQUENCE 84 AA; 9286 MW; 65ED5FE4F44A0B0 CRC64;
 CC
 CC Query Match 4.2%; Score 6; DB 1; Length 84;
 CC Best Local Similarity 100.0%; Pred. No. 33;
 CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 3 DTNPFI 8
 CC DB 60 DTNPFI 65
 CC
 CC RESULT 15
 CC COXD_RAT
 CC ID COXD_RAT STANDARD; PRT; 94 AA.
 CC AC P10817;
 CC DT 01-JUL-1989 (Rel. 11, Created)
 CC DT 01-FEB-1996 (Rel. 33, Last sequence update)
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
 CC DE Cytochrome c oxidase polypeptide VIA-heart, mitochondrial precursor
 CC DE (EC 1.9.3.1) (COXVIA) (Fragment).
 CC GN COX6A2.
 CC OS Rattus norvegicus (Rat).
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 CC OC NCBI_TaxID=10116;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC TISSUE=Heart;
 CC RX MEDLINE=89052650; PubMed=2461293;
 RA Schlierf A., Droste M., Winter M., Kadenbach B.,
 RT "Characterization of two different genes (CDNA) for cytochrome c
 RT oxidase subunit VIA from heart and liver of the rat.",
 RL EMBO J. 7:2387-2391(1988).
 CC [2]
 CC RN [2]
 CC RP REVISIONS.
 CC RC TISSUE=Heart;
 CC RA Kadenbach B.;
 RL Submitted (JUN-1989) to the EMBL/GenBank/DBJ databases.
 CC [3]

SEQUENCE OF 10-19.

RP TISSUE=Heart;

RX MEDLINE=9012894; PubMed=2153407;

RA Kadenbach B., Stroth A., Becker A., Eckersorn C., Lottspeich F.;

RT "Tissue- and species-specific expression of cytochrome c oxidase

RL isozymes in vertebrates.";

CC Biochim. Biophys. Acta 1015:368-372(1990).

CC -!- FUNCTION: THIS PROTEIN IS ONE OF THE NUCLEAR-CODED POLYPEPTIDE

CC CHAINS OF CYTOCHROME C OXIDASE. THE TERMINAL OXIDASE IN

CC MITOCHONDRIAL ELECTRON TRANSPORT.

CC -!- CATALYTIC ACTIVITY: 4 ferrocycytochrome c + O(2) = 4 ferricytochrome

CC c + 2 H(2)O.

CC -!- SUBCELLULAR LOCATION: Mitochondrial inner membrane.

CC -!- TISSUE SPECIFICITY: HEART/MUSCLE SPECIFIC ISOFORM.

CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE VIA FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: X12554; GAA31068.1; ALT_INIT.

DR InterPro: IPR001349; COX6A.

DR Pfam: PF02046; COX6A; 1.

DR ProDom: PDOC6036; COX6A; 1.

DR PROSITE: PS01329; COX6A; 1.

KW Oxidoreductase; Inner membrane; Mitochondrion; Transit peptide.

FT NON_TER 1

FT TRANSIT <1 9 MITOCHONDRION.

FT CHAIN 10 94 CYTOCHROME C OXIDASE POLYPEPTIDE VIA-

FT HEART.

SO SEQUENCE 94 AA; 10487 MW; AAF44BDDACEF50A6 CRC64;

Query Match 4.2%; Score 6; DB 1; Length 94;

Best Local Similarity 100.0%; Pred.No. 36;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 SVALCS 67

DB 36 SVALCS 41

Search completed: November 13, 2003, 13:27:51

Job time : 18 secs

OM protein - protein search, using sw model

Run on: November 13, 2003, 13:25:51 ; Search time 20 Seconds
(without alignments)
687.606 Million cell updates/sec

Title: US-09-868-025-2
Perfect score: 143
Sequence: 1 MVDNFIQSESETHAWCS.....LGVLTTPKMEFAAIAAGKV 143

Scoring table: OLIGO 60.0 , Gapext 60.0
Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR 76.*

1: Pir1.*

2: Pir2.*

3: Pir3.*

4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	5.6	295	2	hypothetical prote
2	8	5.6	623	2	vacuolar sorting r
3	8	5.6	624	2	vacuolar sorting r
4	7	4.9	64	2	hypothetical prote
5	7	4.9	139	1	conserved hypothet
6	7	4.9	205	2	guanylate kinase T
7	7	4.9	243	2	probable transcrip
8	7	4.9	249	2	hypothetical prote
9	7	4.9	280	2	phosphatidate cycl
10	7	4.9	303	2	hypothetical prote
11	7	4.9	318	2	probable sugar upt
12	7	4.9	355	2	branched-chain fat
13	7	4.9	355	2	branched-chain fat
14	7	4.9	359	2	B-cell surface ant
15	7	4.9	377	2	hypothetical prote
16	7	4.9	386	2	probable fadA6 pro
17	7	4.9	398	2	yeam protein - Esc
18	7	4.9	398	2	1-deoxy-D-xylose
19	7	4.9	398	2	1-deoxy-D-xylose
20	7	4.9	398	2	1-deoxy-D-xylose
21	7	4.9	398	2	1-deoxy-D-xylose
22	7	4.9	437	2	agave protein (impo
23	7	4.9	550	2	hypothetical prote
24	7	4.9	597	2	protein kinase C3
25	7	4.9	882	2	probable peptidogl
26	7	4.9	903	2	probable peptidogl
27	7	4.9	1101	2	probable ATP-depen
28	7	4.9	1274	2	probable membrane
29	7	4.9	1829	2	probable sensory h

30 6 4.2 65 2 AC0779
31 6 4.2 85 2 S01157
32 6 4.2 87 2 G75507
33 6 4.2 88 2 T29334
34 6 4.2 96 2 T37076
35 6 4.2 97 2 C82257
36 6 4.2 101 2 D97386
37 6 4.2 101 2 AD2604
38 6 4.2 105 2 E41377
39 6 4.2 114 2 AP3277
40 6 4.2 117 2 C83576
41 6 4.2 121 2 AH3396
42 6 4.2 125 2 G87491
43 6 4.2 127 2 H73339
44 6 4.2 128 2 E88449
45 6 4.2 130 2 I59298

ALIGNMENTS

RESULT 1

hypothetical protein alr2373 [imported] - Nostoc sp. (strain PCC 7120)
C/Species: Nostoc sp. PCC 7120
A/Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C/Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C/Accession: AF2102
R/Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguc
DNA Res. 8, 205-213, 2001
A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium A
A/Reference number: AB1807, MUID:21595285; PMID:11759840
A/Accession: AF2102
A/Status: Preliminary
A/Molecule type: DNA
A/Residues: 1-295 <KUR>
A/Cross-references: GB:BA000019; PIDN:BA074072.1; PID:gi7113465; GSPOB:GN00179
A/Experimental source: strain PCC 7120
C/Genetics:
A/Genes: alr2373
C/Superfamily: Streptomyces peucetius daunorubicin resistance protein

Query Match 5.6%; Score 8; DB 2; Length 295;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 40 LALGVTAI 47
|||
DB 178 LALGVTAI 185
|||
RESULT 2
T06794
vacuolar sorting receptor protein BP-80 - garden pea
C/Species: Pisum sativum (garden pea)
C/Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 11-May-2000
C/Accession: T06794
R/Paris, N.; Rogers, S.W.; Jiang, L.; Kirsch, T.; Beevers, L.; Phillips, T.E.; Rogers
Plant Physiol. 115, 29-39, 1997
A/Title: Molecular cloning and further characterization of a probable plant vacuolar
A/Reference number: Z15236; MUID:97451769; PMID:9306690
A/Accession: T06794
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-623 <PAR>
A/Cross-references: EMBL:U79958; NID:gi737221; PIDN:AA072110.1; PID:gi737222

Query Match 5.6%; Score 8; DB 2; Length 623;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 81 CLTNGGC 88

b 415 CLTNGGC 422
|||||
acicular sorting receptor protein homolog PV72 - cucurbit
;Species: Cucurbita sp. (cucurbit)
;Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 21-Jul-2000
;Accession: T00044
;Shimada, T.; Kuroyanagi, M.; Nishimura, M.; Hara-Nishimura, I.
;Title: A pumpkin 72-kDa membrane protein of precursor-accumulating vesicles has charac
;Reference number: 214076; MUID:98182943; PMID:9522472
;Accession: T00044
;Status: translated from GB/EMBL/DBJ
;Molecule type: mRNA
;Residues: 1-624 <SH1>
;Cross-references: EMBL:AB006809; NID:92943791; PIDN:BA025079.1; PID:92943792
;Experimental source: subsp. Kurokawa Amakuri Nankin, cotyledon
Query Match 5.6%; Score 8; DB 2; Length 624;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y 81 CLTNGGC 88
|||||
b 417 CLTNGGC 424
|||||
RESULT 4
X0648
;Protein names: ORF3 protein
;Species: Streptomyces lividans
;Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 22-Oct-1999
;Accession: P06648
;Sedlmeier, R.; Linti, G.; Gregor, K.; Schmieger, H.
;Title: Sequences of tRNA-encoding genes and associated open reading frames of Streptom
;Reference number: JN0849; MUID:94010335; PMID:7691688
;Accession: P06648
;Molecule type: DNA
;Residues: 1-64 <SED>
;Cross-references: EMBL:X65556; NID:947167; PIDN:CAA46527.1; PID:947170
Query Match 4.9%; Score 7; DB 2; Length 64;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y 40 LALGVTA 46
|||||
b 34 LALGVTA 40
|||||
RESULT 5
69280
;Protein names: conserved hypothetical protein AF0241 - Archaeoglobus fulgidus
;Species: Archaeoglobus fulgidus
;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
;Accession: A69280
;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
;Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
;Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae
;Reference number: 364-370, 1997
;Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artlich, P.; Kaine, B.P.; Sykes, S.
;Smith, H.O.; Woese, C.R.; Venter, J.C.
;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae
;Reference number: A69280
;Accession: A69280
;Status: preliminary; nucleic acid sequence not shown; translation not shown
;Molecule type: DNA
;Residues: 1-139 <KLE>
A;Cross-references: GB:AE001089; GB:AE000782; NID:92689412; PIDN:AAB90992.1; PID:92650400
C;Superfamily: conserved hypothetical protein MJ1583
Query Match 4.9%; Score 7; DB 1; Length 139;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 23 TTRSPSR 29
|||||
DB 80 TTRSPSR 86
|||||
RESULT 6
P81718
;Protein names: guanylate kinase TC0299 [imported] - Chlamydia muridarum (strain Nigg)
C;Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Jun-2000
C;Accession: P81718
R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, I.
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A;Reference number: A81500; MUID:20150255; PMID:10684935
A;Accession: P81718
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-205 <TET>
A;Cross-references: GB:AE002297; GB:AE002160; NID:97190334; PIDN:AAF39164.1; PID:97190334
A;Experimental source: strain Nigg (MoPn)
C;Genetics:
X;Gene: TC0299
C;Superfamily: guanylate kinase; guanylate kinase homology
Query Match 4.9%; Score 7; DB 2; Length 205;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 21 SSITRSP 27
|||||
DB 51 SSITRSP 57
|||||
RESULT 7
B95924
;Protein names: probable transcription regulator protein (imported) - Sinorhizobium meliloti (strain 102)
C;Species: Sinorhizobium meliloti
C;Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C;Accession: B95924
R;Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, P.J.; Hernan
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A;Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo
A;Reference number: A95842; MUID:21396508; PMID:11481431
A;Accession: B95924
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-243 <KUR>
A;Cross-references: GB:AL591985; PIDN:CAC49058.1; PID:915140543; GSPDB:GN00167
A;Experimental source: strain 1021, megaplasmid pSymB
R;Galibert, P.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, P.; Barloy-Hubler,
Pala, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebaull, P.; Vandenbol, M.; Vorholter, P.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yen, K.C
A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A;Reference number: A96039; MUID:21368234; PMID:11474104
A;Contents: annotation
C;Genetics:
A;Gene: SMB21079
A;Genome: plasmid
Query Match 4.9%; Score 7; DB 2; Length 243;
Best Local Similarity 100.0%; Pred. No. 17;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 ALGVTAI 47
 |||||
 Db 190 ALGVTAI 196

RESULT 8
 T25643
 hypothetical protein C46H11.9 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000
 C:Accession: T25643
 R:Miller, N.; Bradshaw, H.; Wamsley, P.
 submitted to the EMBL Data Library, February 1997
 A:Description: The sequence of C. elegans cosmid C46H11.
 A:Reference number: Z20061
 A:Accession: T25643
 A:Status: preliminary;
 A:Molecule type: DNA
 A:Residues: 1-249 <ML>
 A:Cross-references: EMBL:U8314; PIDN:AA842363.1; GSPDB:GN00019; CESP:C46H11.9
 A:Experimental source: strain Bristol N2, clone C46H11
 C:Genetics:
 A:Gene: CESP:C46H11.9
 A:Map position: 1
 A:Introns: 28/1; 55/1; 205/1
 C:Superfamily: Caenorhabditis elegans hypothetical protein F48G7.8

Query Match 4.9%; Score 7; DB 2; Length 249;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 SSSSTRS 26
 |||||
 Db 191 SSSSTRS 197

RESULT 9
 E82099
 phosphatidate cytidyltransferase VC2255 [imported] - Vibrio cholerae (strain N16961 se
 C:Species: Vibrio cholerae
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
 C:Accession: E82099
 R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
 Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.
 1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
 Nature 406, 477-483, 2000
 A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
 A:Reference number: A82035; MUID:20406833; PMID:10952301
 A:Accession: E82099
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-280 <HA>
 A:Cross-references: GB:AE004297; GB:AE003852; MID:99556810; PIDN:AAF95399.1; GSPDB:GN001
 A:Experimental source: serogroup O1; strain N16961; biotype El Tor
 C:Genetics:
 A:Gene: VC2255
 A:Map position: 1
 C:Superfamily: phosphatidate cytidyltransferase

Query Match 4.9%; Score 7; DB 2; Length 280;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 GGISVAL 65
 |||||
 Db 62 GGISVAL 68

RESULT 10
 D98210
 hypothetical protein AGR_L_1257 [imported] - Agrobacterium tumefaciens (strain C58, Cerd

C:Species: Agrobacterium tumefaciens
 C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
 C:Accession: D98210
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldma
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, I
 Science 294, 2323-2328, 2001
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium t
 A:Reference number: A97359; MUID:21608551; PMID:11743194
 A:Accession: D98210
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-303 <KUR>
 A:Cross-references: GB:AE007870; PIDN:AAK89206.1; PID:915159026; GSPDB:GN00170
 C:Genetics:
 A:Gene: AGR_L_1257
 A:Map position: linear chromosome

Query Match 4.9%; Score 7; DB 2; Length 303;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 LALGVTA 46
 |||||
 Db 126 LALGVTA 132

RESULT 11
 E96033
 probable sugar uptake ABC transporter permease protein SMD20621 [imported] - Sinorhiz
 C:Species: Sinorhizobium meliloti
 C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
 C:Accession: E96033
 R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Herr
 Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
 A:Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing er
 A:Reference number: A95842; MUID:21396508; PMID:11481431
 A:Accession: E96033
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-318 <KUR>
 A:Cross-references: GB:AL591985; PIDN:CAC49933.1; PID:915141421; GSPDB:GN00167
 A:Experimental source: strain 1021, megaplasmid pSymB
 R:Galbert, P.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Ruble
 pella, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F
 L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Kemp, C.; Lelau
 hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh,
 A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
 A:Reference number: A96039; MUID:21368234; PMID:11474104
 A:Contents: annotation
 C:Genetics:
 A:Gene: SMD20621
 A:Genome: Plasmid
 C:Superfamily: l-arabinose transport system permease araH

Query Match 4.9%; Score 7; DB 2; Length 318;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 LALGVTA 46
 |||||
 Db 36 LALGVTA 42

RESULT 12
 AB1246
 branched-chain fatty-acid kinase homolog lmo1370 [imported] - Listeria monocytogenes (

C:Species: Listeria monocytogenes
 C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 17-May-2002
 C:Accession: AB1246
 R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Anand, A.; Baquero, F.; Berche, P.; Bloech
 ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Feith,

D.; Jones, L.M.; Karst, U.
 Science 294, 849-852, 2001
 A:Authors: Kref, J.; Kuhn, M.; Kunat, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
 Ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
 A:Title: Comparative genomics of *Listeria* species
 A:Reference number: AB1077; MUID:21337279; PMID:11679669
 A:Accession: AB1246
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-355 <GLA>
 A:Cross-references: GB:NC_003210; PIDN:CAC99448.1; PID:G16410786; GSPDB:GN00177
 A:Experimental source: strain EGD-e
 C:Genetics:
 A:Gene: lml370
 C:Superfamily: acetate kinase

Query Match 4.9%; Score 7; DB 2; Length 355;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 GGGISVA 64
 DB 185 GGGISVA 191

RESULT 13
 AF1608
 branched-chain fatty-acid kinase homolog lml407 [imported] - *Listeria innocua* (strain O
 C:Species: *Listeria innocua*
 C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 17-May-2002
 A:Accession: AF1608
 R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, P.; Berche, P.; Bloeker
 R.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
 D.; Jones, L.M.; Karst, U.
 Science 294, 849-852, 2001
 A:Authors: Kref, J.; Kuhn, M.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
 Ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
 A:Title: Comparative genomics of *Listeria* species
 A:Reference number: AB1077; MUID:21337279; PMID:11679669
 A:Accession: AF1608
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-355 <GLA>
 A:Cross-references: GB:AL592023; PIDN:CAC96638.1; PID:G16413880; GSPDB:GN00178
 A:Experimental source: strain Clp11262
 C:Genetics:
 A:Gene: lml407
 C:Superfamily: acetate kinase

Query Match 4.9%; Score 7; DB 2; Length 355;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 GGGISVA 64
 DB 185 GGGISVA 191

RESULT 14
 A4352
 B-cell surface antigen CD72 - human
 N:Alternate names: B-cell differentiation antigen Lyb-2 homolog
 C:Species: *Homo sapiens* (man)
 C:Date: 28-Oct-1992 #sequence_revision 28-Oct-1992 #text_change 21-Jul-2000
 C:Accession: A4352
 R:van Hoegen, I.; Nakayama, E.; Parnes, J.R.
 J. Immunol. 144, 4870-4877, 1990
 A:Title: Identification of a human protein homologous to the mouse Lyb-2 B cell differer
 A:Reference number: A4352; MUID:90278102; PMID:2141045
 A:Accession: A4352
 A:Molecule type: mRNA
 A:Residues: 1-359 <VON>
 A:Cross-references: GB:M54992; NID:G187262; PIDN:AAA36189.1; PID:G187263

C:Genetics:
 A:Gene: GDB:CD72
 A:Cross-references: GDB:128153; OMIM:107272
 A:Map position: 9p-9p
 C:Superfamily: C-type lectin homology
 C:Keywords: B-cell; glycoprotein; homodimer; surface antigen; transmembrane protein
 E:1-95/Domain: intracellular #status predicted <CVT>
 E:96-116/Domain: transmembrane #status predicted <TMM>
 F:117-359/Domain: extracellular #status predicted <EXT>
 F:136/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 4.9%; Score 7; DB 2; Length 359;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 LGVTAIC 48
 DB 107 LGVTAIC 113

RESULT 15
 AD1057
 hypothetical protein STY4784 [imported] - *Salmonella enterica* subsp. *enterica* serovar
 C:Species: *Salmonella enterica* subsp. *enterica* serovar Typhi
 A:Note: this species has also been called *Salmonella typhi*
 C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
 C:Accession: AD1057
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church
 th, T.; Conner, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farra
 S.; Moule, S.; O'Gaora, P.
 Nature 413, 848-852, 2001
 A:Authors: Park, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K
 A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* ser
 A:Reference number: AB0502; MUID:21534947; PMID:11677608
 A:Accession: AD1057
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-377 <PAR>
 A:Cross-references: GB:AL513382; PIDN:CAD06905.1; PID:G16505553; GSPDB:GN00176
 C:Genetics:
 A:Gene: STY4784

Query Match 4.9%; Score 7; DB 2; Length 377;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 136 RAIRAGK 142
 DB 356 RAIRAGK 362

Search completed: November 13, 2003, 13:29:13
 Job time : 22 secs

GenCore version 5.1.6
Copyright. (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 14, 2003, 12:25:29 ; Search time 57 Seconds
(without alignments)
1107.330 Million call updates/sec

Title: US-09-868-025-2
Perfect score: 803
Sequence: 1 MVDNFFISEQSETHAWCWS.....LGVLTTPHKKPFAAIRAGKV 143

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -p2n.model -DEV=xlh
-MODE=frame_p2n.model/US09868025/runat 13112003 130406 17492/app query.fasta_1.327
-Q=/cg2_1/USPTO_spool/US09868025/SUFFIX=nrni -MINMATCH=0.1 -LOOFC=0
-DB=Issued Patents_NA -QFMT=tastep -SUFFIX=nrni -MINMATCH=0.1 -LOOFC=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MAIRIX=BLOSUM62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09868025 @CGN 1 1.44 @runat 13112003 130406 17492 -NCPU=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELEX=7

Database : Issued Patents_NA:
1: /cg2_6/prodata/2/ina/5A-COMB.seq:
2: /cg2_6/prodata/2/ina/5B-COMB.seq:
3: /cg2_6/prodata/2/ina/6A-COMB.seq:
4: /cg2_6/prodata/2/ina/6B-COMB.seq:
5: /cg2_6/prodata/2/ina/6C-COMB.seq:
6: /cg2_6/prodata/2/ina/backfiles1.seq:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	89.5	11.1	1092	3	US-09-077-675A-15
2	89.5	11.1	1092	4	US-09-077-674-15
3	84	10.5	2043	3	US-08-191-160-20
4	84	10.5	3750	3	US-08-191-160-22
C 5	83.5	10.4	4403765	3	US-09-103-840A-2
C 6	83.5	10.4	4411529	3	US-09-103-840A-1
7	82.5	10.3	36241	4	US-08-311-731A-134
8	81	10.1	31571	1	US-08-323-443B-1
9	81	10.1	53526	3	US-08-558-136-2
10	81	10.1	53577	3	US-08-558-136-1
11	80	10.0	5648	3	US-09-371-008-1
C 12	79.5	9.9	2043	4	US-09-252-991A-13911

ALIGNMENTS

RESULT 1
US-09-077-675A-15
; Sequence 15, Application US/09077675A
; Patent No. 6242199
; GENERAL INFORMATION:
; APPLICANT: Pai Lee-Yuh
; APPLICANT: Feighner, Scott C.
; APPLICANT: Howard, Andrew D.
; APPLICANT: Pong, Sheng-Shung
; APPLICANT: Van Der Ploeg, Leonardus H.T.
; TITLE OF INVENTION: RECEPTOR ASSAY
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: P.O. Box 2000, 126 E. Lincoln Ave.
; CITY: Rahway
; STATE: NJ
; COUNTRY: USA
; ZIP: 07065-0900
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/077,675A
; FILING DATE: 3-JUN-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Cocuzzo, Anna L.
; REGISTRATION NUMBER: 42,452
; REFERENCE/DOCKET NUMBER: 19590P
; TELECOMMUNICATION INFORMATION:

C 13	79	9.8	1083	2	US-08-119-866-1	Sequence 1, Appl
C 14	79	9.8	1139	3	US-09-124-758-5	Sequence 5, Appl
C 15	79	9.8	1139	4	US-09-709-677-5	Sequence 5, Appl
C 16	79	9.8	2287	4	US-09-220-132-15	Sequence 15, Appl
C 17	79	9.8	11531	1	US-08-068-945A-1	Sequence 1, Appl
C 18	79	9.8	11531	1	US-08-442-806-1	Sequence 1, Appl
C 19	79	9.8	11531	4	US-09-355-295B-1	Sequence 1, Appl
20	78.5	9.8	1357	3	US-09-383-586-8	Sequence 8, Appl
21	78.5	9.8	17410	1	US-07-841-626-3	Sequence 3, Appl
22	78.5	9.8	17410	1	US-08-147-043-3	Sequence 3, Appl
23	78.5	9.8	17410	1	US-08-447-570-3	Sequence 3, Appl
24	78.5	9.8	17410	2	US-08-449-700-3	Sequence 3, Appl
25	78.5	9.8	17410	2	US-08-449-699A-3	Sequence 3, Appl
26	78.5	9.8	17410	4	US-09-148-325C-3	Sequence 3, Appl
27	78.5	9.8	17410	3	US-08-957-425-3	Sequence 3, Appl
28	78.5	9.8	17415	3	US-08-486-343A-1	Sequence 1, Appl
29	78.5	9.8	17415	5	PCT-US95-07349-1	Sequence 1, Appl
30	78	9.7	741	1	US-08-324-977-43	Sequence 43, Appl
31	78	9.7	741	2	US-08-384-616-43	Sequence 43, Appl
32	78	9.7	741	3	US-08-904-686A-43	Sequence 43, Appl
33	78	9.7	741	3	US-09-315-850-43	Sequence 1, Appl
34	78	9.7	1617	2	US-08-735-041A-1	Sequence 1, Appl
35	78	9.7	1617	3	US-09-190-476B-1	Sequence 1, Appl
36	78	9.7	1617	3	US-09-190-889A-1	Sequence 1, Appl
37	78	9.7	1617	3	US-09-190-938B-1	Sequence 2, Appl
38	78	9.7	1934	4	US-08-653-648A-2	Sequence 2, Appl
39	78	9.7	1934	4	US-08-653-648A-9	Sequence 9, Appl
40	78	9.7	2389	1	US-08-123-161A-13	Sequence 13, Appl
41	78	9.7	2389	1	US-08-483-278-13	Sequence 13, Appl
42	78	9.7	2463	4	US-08-653-648A-10	Sequence 10, Appl
43	78	9.7	2464	4	US-08-653-648A-3	Sequence 3, Appl
44	78	9.7	3171	5	PCT-US95-09261-1	Sequence 1, Appl
45	78	9.7	6039	1	US-08-324-977-11	Sequence 11, Appl

TELEPHONE: 732-594-1273
TELEFAX: 732-594-4720
TELEX:
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 1092 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-077-675A-15

Alignment Scores:
Pred. No.: 1.88 Length: 1092
Score: 89.50 Matches: 38
Percent Similarity: 35.21% Conservatives: 12
Best Local Similarity: 26.76% Mismatches: 50
Query Match: 11.15% Indels: 42
DB: 3 Gaps: 7

US-09-868-025-2 (1-143) x US-09-077-675A-15 (1-1092)

QY 17 TPCysrTpSerSerThrThr-ArgSerProSerArgHisHisLeuHisArgGluar 36
DB 542 TGGCGTGGAGCAGCAACGACAGATCCCGGGACACCAAGATGCGGCGCCACCG 601
QY 36 gileProCysLeuAlaLeuGlyValThrAlaIleCysSerLeuValTrpIleHisValse 56
DB 602 AGTTCGCTGT-----GCGCTCTGGGCTCTCTCACCG 631
QY 56 rHisGlyGlyGlyLe-----SerValAlaLeuCysSerGlnCys--Le 70
DB 632 TCATGTGTGGTGTCTCAGGCTTCTTCTTCTACCGCTCTTCTGCTCTCTCTCT 691
QY 70 uGlnThrAsn-----Al 74
DB 692 ACAGTCTCATCGGAGGAGCTATGCGGAGACCGGAGATGCGGCGCTCGCCTCGC 751
QY 74 aLeuArgProArgProArgCysLeuThrAsnAsnGlyCysTyrrGlyGluCysHisG1 94
DB 752 TCCGGACCCAGAACACACAGACAGATGCTGTGTGTGTGTGTGTGTGTGTTC 811
QY 94 ySerLeu-----GlyHisValAspArgPheProGlnHisSerAsnG1 108
DB 812 TCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTTTC 868
QY 108 uTrpAsnSerGlyMetAspSerCysLysProLeuArgGlyGluPheLeuGlyValLeuth 128
DB 869 CTGGCTCTCTGGAGATGCTCAGATCAGCCAGTACTGCAACCT-----GGTGTCTCTTG 922
QY 128 rPro 129
DB 923 TCCT 926

RESULT 2

US-09-077-674-15
Sequence 15, Application US/09077674
Patent No. 6531314
GENERAL INFORMATION:
APPLICANT: Arena, Joseph P.
APPLICANT: Cully, Doris F.
APPLICANT: Feighner, Scott D.
APPLICANT: Howard, Andrew D.
APPLICANT: Liberator, Paul A.
APPLICANT: Schaeffer, James M.
APPLICANT: Van Der Ploeg, Leonardus
TITLE OF INVENTION: GROWTH HORMONE SECRETAGOGUE RECEPTOR FAMILY
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merck & Co., Inc.
STREET: P.O. Box 2000, 126 E. Lincoln Ave.
CITY: Rahway
STATE: NJ

COUNTRY: USA
ZIP: 07065-0900
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/077,674
FILING DATE: 3-JUN-1998
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Cocuzzo, Anna L.
REGISTRATION NUMBER: 42,452
REFERENCE/DOCKET NUMBER: 195899
TELECOMMUNICATION INFORMATION:
TELEPHONE: 732-594-1273
TELEFAX: 732-594-4720
TELEX:
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 1092 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-077-674-15

Alignment Scores:
Pred. No.: 1.88 Length: 1092
Score: 89.50 Matches: 38
Percent Similarity: 35.21% Conservatives: 12
Best Local Similarity: 26.76% Mismatches: 50
Query Match: 11.15% Indels: 42
DB: 3 Gaps: 7

US-09-868-025-2 (1-143) x US-09-077-674-15 (1-1092)

QY 17 TPCysrTpSerSerThrThr-ArgSerProSerArgHisHisLeuHisArgGluar 36
DB 542 TGGCGTGGAGCAGCAACGACAGATCCCGGGACACCAAGATGCGGCGCCACCG 601
QY 36 gileProCysLeuAlaLeuGlyValThrAlaIleCysSerLeuValTrpIleHisValse 56
DB 602 AGTTCGCTGT-----GCGCTCTGGGCTCTCTCACCG 631
QY 56 rHisGlyGlyGlyLe-----SerValAlaLeuCysSerGlnCys--Le 70
DB 632 TCATGTGTGGTGTCTCAGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 691
QY 70 uGlnThrAsn-----Al 74
DB 692 ACAGTCTCATCGGAGGAGCTATGCGGAGACCGGAGATGCGGCGCTCGCCTCGC 751
QY 74 aLeuArgProArgProArgCysLeuThrAsnAsnGlyCysTyrrGlyGluCysHisG1 94
DB 752 TCCGGACCCAGAACACACAGACAGATGCTGTGTGTGTGTGTGTGTGTGTTC 811
QY 94 ySerLeu-----GlyHisValAspArgPheProGlnHisSerAsnG1 108
DB 812 TCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTTTC 868
QY 108 uTrpAsnSerGlyMetAspSerCysLysProLeuArgGlyGluPheLeuGlyValLeuth 128
DB 869 CTGGCTCTCTGGAGATGCTCAGATCAGCCAGTACTGCAACCT-----GGTGTCTCTTG 922
QY 128 rPro 129
DB 923 TCCT 926

QY	15	HISALATrPCysA-----TtpSerSerThrThraArgSerProSerArgHishisLeu 32
Db	194	CACCCTTCGTGATAGGCTGGAGCCGTCAAAACGAGGTGCACCTCACA-CACCCCATA 252
QY	33	HISARGGIUAArgileProCysLeuAlaLeuGlyValThrAlaIleCysSerLeuValTrp 52
Db	253	ACCAAAATTTCATCGGCATGCATGCAGCCACCTGGAGGTGCGTCACGAGC---ACCTGG 309
QY	53	IleHisValSerHisGlyGlyIleSerValAlaLeuCysSerGlnCysLeuGlnThr 72
Db	310	GTGCTGGTG-----GGCGGGGTCTTCACACTCTGGCTGGCTATTGTTTACAACA 360
QY	73	AenAla-----LeuArgProArgProAspCysLeuThrAsn 84
Db	361	GGCAGCGTGGTCATTGTGGTAGGATCATCTTGTCCGGCGCGCGGCTATTGTTCCCGAC 420
QY	85	AsnGlyGlyCysTyr-----GlyGluCysHisGlySerLeuGlyHis 98
Db	421	ACGGAAGTCCTTACACAGAGTTCGATGATGAGAGTGGCGGTGCCACTCCCTTAC 480
QY	99	ValAspArgPheProGlnHisSerAsnGluTrpAsnSerGlyMetAspSerCysLysPro 118
Db	481	ATCGAGCAG-----GGAATGCAGCTCGCCGAGCAG 510
QY	119	LeuArgGlyGluPheLeuGlyValLeuThrProHisProLysMetGluPheAlaAla 137
Db	511	TTCRAGCAAAAGCGCTGGGTGTGTCAGACAGCCACCAAGCAGCGGCGCT 567
 RESULT 4 US-08-191-160-22 ; Sequence 22, Application US/08191160 ; Patent No. 6210675 ; GENERAL INFORMATION: ; APPLICANT: Highfield, Peter Edmund ; APPLICANT: Rodgers, Brian Colin ; APPLICANT: Tedder, Richard Seton ; APPLICANT: Barbara, John Anthony James ; TITLE OF INVENTION: Viral Agent ; NUMBER OF SEQUENCES: 25 ; CORRESPONDENCE ADDRESS: ; ADDRESSEE: Rothwell, Figg, Ernst & Kurz ; STREET: 1700 K Street ; CITY: Washington ; STATE: D.C. ; COUNTRY: U.S.A. ; ZIP: 20006 ; COMPUTER READABLE FORM: ; MEDIUM TYPE: Floppy diskette, 5.25 inch, 360 Kb storage ; COMPUTER: IBM AT compatible ; OPERATING SYSTEM: MS-DOS V3.2 ; SOFTWARE: Wordperfect 5.0 (DOS text) ; CURRENT APPLICATION DATA: ; APPLICATION NUMBER: US/08/191,160 ; FILING DATE: ; CLASSIFICATION: ; PRIORITY APPLICATION DATA: ; APPLICATION NUMBER: 07/628,516 ; FILING DATE: 17 DEC 1990 ; APPLICATION NUMBER: UK 89 28 562.1 ; FILING DATE: 18 DEC 1989 ; PRIOR APPLICATION DATA: ; APPLICATION NUMBER: UK 90 04 414.0 ; FILING DATE: 27 FEB 1990 ; PRIOR APPLICATION DATA: ; APPLICATION NUMBER: UK 90 04 814.1 ; FILING DATE: 03 MAR 1990 ; ATTORNEY/AGENT INFORMATION: ; NAME: E. Anthony Figg ; REGISTRATION NUMBER: 27,195 ; REFERENCE/DOCKET NUMBER: 1645-103A ; TELECOMMUNICATION INFORMATION: ; TELEPHONE: (202) 833-5740 ; TELEFAX: (202) 833-5744		
 Alignment Scores: Pred. No.: 17.1 Length: 2043 Score: 84.00 Matches: 36 Percent Similarity: 41.01% Conservative: 21 Best Local Similarity: 25.90% Mismatches: 52 Query Match: 10.46% Indels: 31 DB: Gaps: 6 US-09-868-025-2 (1-143) x US-08-191-160-20 (1-2043)		

US-09-868-025-2 (1-143) x US-08-191-160-20 (1-2043)

Best Local Similarity: 25.9% Mismatches: 47
Query Match: 10.4% Indels: 36
DB: 3 Gaps: 6
US-09-868-025-2 (1-143) x US-09-103-840A-1 (1-4411529)
Qy 4 ThrAsnPheProIleSerGluGlnSerGluThrHisAlaTrp----- 17
Db 1258971 TCGAACTCCCGGTCGCGCGGCTACCAACCGGTACTCTGGTTTGGCGATGATGGC 1258912
Qy 18 CysTrpSerSerThrThrArgSerProSerArgHisHisLeuHisArgGluArgIle 37
Db 1258911 TCGCGTCGAGCAGCACCCCGACCTCACCGGTTGGAGCCGACCGC----- 1258861
Qy 38 ProCysLeuAlaLeuGlyValThrAlaIleCysSerLeuValTrpIleHisValSerH1 57
Db 1258860 ---TGTTCATCGCCGCGATCAGTGTCTGTTGTGCA----- 1258827
Qy 57 GelyGlyIleSerValAlaLeuCysSerGlnCysLeu----- 70
Db 1258826 -GGTGGCGTCGCCGAAGTTCAGCGCGAGTGTCTGACACCGGTTCGCGGTGCGT 1258768
Qy 71 -----GlnThrAsnAlaLeuArgProArgProArgCysLeuThrAsnGlyG1 87
Db 1258767 ATGACGTAGTCTGTCTTCGTATGCGCGGCGCATCGCCGATGCG-----CCAGC 1258720
Qy 87 yCysTyrglyGlyCysHisGlySerLeuGlyHisValAspArgPheProGlnHisSerAs 107
Db 1258719 GTAATAGCGCCCTGCTCATGGAGTTCATCGCACGAACCGTATCCAGATACATCG-- 1258662
Qy 107 nGluTrpAsnSerGlyMetAspSerCysLys 117
Db 1258661 -----GGATCGGCTCGTCTTGGCGT 1258638
RESULT 7
US-08-311-731A-134
Sequence 134, Application US/08311731A
Patent No. 5683266
GENERAL INFORMATION:
APPLICANT: SMITH, DOUGLAS
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 411
CORRESPONDENCE ADDRESS:
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/311,731A
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: C0044/7125
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/720-3500
TELEFAX: 617/720-2441
INFORMATION FOR SEQ ID NO: 134:
SEQUENCE CHARACTERISTICS:
LENGTH: 36241 base pairs
TYPE: nucleic acid
STRANDEDNESS: double

TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Mycobacterium leprae
US-08-311-731A-134
Alignment Scores:
Pred. No.: 1,3e+03 Length: 36241
Score: 82.50 Matches: 42
Percent Similarity: 35.09% Conservative: 18
Best Local Similarity: 24.56% Mismatches: 57
Query Match: 10.27% Indels: 54
DB: 4 Gaps: 9
US-09-868-025-2 (1-143) x US-08-311-731A-134 (1-36241)
Qy 18 CysTrpSerSerThrThrArgSerProSerArgHisHisLeuHisArgGluArg--- 36
Db 35134 TGTATCTCTGAGAAATTCGCGCGCGCAACCAATTTCCACCTGAGCGGACACGCTAG 35193
Qy 37 -----IleProCysLeuAlaLeuGlyValThrAlaIleCys----- 48
Db 35194 AACGTAAGTTTCCGAACAACCGGTGCTTGGCCTTG-----ATCGCCTGCTGTATTG 35247
Qy 49 -----SerLeuValTrp-----Ile 53
Db 35248 ATCGCGACGTCCCTCGTGTGTGTAATGCGTAATGCGAGCGATGGCGAGCAGCGCACC 35307
Qy 54 HisValSerHisGlyGlyIleSerValAlaLeuCysSerGlnCysLeuGlnThrAsn 73
Db 35308 CACACCAAGCAGCAGCGCGCGGTTGATCGTGTGTAGACAATGTGCTGCGCATCAC 35367
Qy 74 AlaLeu---Arg-ProArgProAspCysLeuThrAsn---As 85
Db 35368 CAATTCACGCGCCCGCCGACACATCGCGCAACGTGCGATCACTATCTCAAGAAA 35427
Qy 85 nGlyGlyCysTyrglyGlyCysHisGlySerLeuGlyHisVal----- 99
Db 35428 CATCAATGCTCTCGGACCGTCCGAGCGACGTCACAACTAGTATTTCGACACCG 35487
Qy 100 ---AspArgPheProGlnHisSerAsnGluTrpAsnSerGlyMetAspSerCysLysPr 118
Db 35488 CTGAGAACAAATTCGTAGTATGGCACATAGCTGGAACACCGAC-----AACCG 35535
Qy 118 OleuArgGlyGluPheLeuGly-----ValLeuThrProH1 130
Db 35536 GTTACCGGGGAACGACTCTCTTTGAGAAAGCAGTCAGCGCGTGAGTATACCCCCCCC 35595
Db 35596 CCGGAAGCGGCGTACGCTACGTTACGAGCT 35626
RESULT 8
US-08-323-443B-1
Sequence 1, Application US/08323443B
Patent No. 5654170
GENERAL INFORMATION:
APPLICANT: KLINGER, KATHERINE W.
APPLICANT: LANDES, GREGORY M.
APPLICANT: BURN, TIMOTHY C.
APPLICANT: CONNORS, TIMOTHY D.
APPLICANT: DACKOWSKI, WILLIAM R.
APPLICANT: GERMINO, GREGORY
APPLICANT: QIAN, FENG
TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby PC
STREET: 805 Third Avenue
CITY: New York
STATE: NY

COUNTRY: USA
 ZIP: 10022
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/323.443B
 FILING DATE: 12-OCT-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Ludwig, S. Peter
 REGISTRATION NUMBER: 25,351
 REFERENCE/DOCKET NUMBER: 0372/0A462
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 527-7700
 TELEFAX: (212) 753-6237
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 31571 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 IMMEDIATE SOURCE:
 CLONE: PKD1 GENOMIC
 US-08-323-443B-1

Alignment Scores:
 Pred. No.: 1,55e+03 Length: 31571
 Score: 81.00 Matches: 37
 Percent Similarity: 37.93% Conservative: 18
 Best Local Similarity: 25.52% Mismatches: 64
 Query Match: 10.09% Indels: 26
 DB: 1 Gaps: 8

US-09-868-025-2 (1-143) x US-08-323-443B-1 (1-31571)

QY 11 GlnSerGluThrHisAlaTrpCysTrpSerSer-----SerThrThrArgSerProSer 28
 DB 11534 GAGGAGACACACATGTTCTTTTCTGGAGCTCTGAGTGGCCACGCGGACCCCGCC 11593

QY 29 ArgHisHisLeuHisArgGluArgIleProCysLeuAlaLeuGlyValThrAlaIleCys 48
 DB 11594 AAGCACCTGGAAATCAAAACATCCCGCTGCTCTGGGCTGGCTGCACTCTGCTGCT 11653

QY 49 SerLeuValTrpIleHisValSerHis-----GlyGlyGlyIleSerVal 63
 DB 11654 GCGCTCCAGCTGGCTGAGCGCGGCGGCTGCTGGGCGCACAGCGGGGGCGCCACAGTC 11713

QY 64 AlaLeuCysSerGlnCysLeuGlnThrAsnAlaLeuArgProArgProAspCysLeuThr 83
 DB 11714 TCCCTGCAGAGTGAGCGGAGCTGGAAATGCACTCAGCTCAGCCCTTCCGAGAACCTCGC 11773

QY 84 AsnAsn-----GlyGlyCysTyrcGlyGlyCysHisGlySerLeuGlyHisValAspArg 101
 DB 11774 TCTTCATGGCTGGCAGCTGT-----CCTTGGCTAGGGGCT-----AGG 11812

QY 102 PheProGlnHis-----SerAsnGluTrpAsn---SerGlyMetAspSer 115
 DB 11813 GTGCCAGGCACTGGTGGCAGGAGAGGCTACATCTGGGCTGAGGGGCTGGGCTCT 11872

QY 116 CysLysProLeuArgGlyGluPheLeuGlyValThrProHisProLysMetGluPhe 135
 DB 11873 TTTCTCCCTGACAGCTCCGAG-----GCCACCGCTGGCCCGACGCTGGCATTC 11920

QY 136 AlaAlaIleArgAla 140
 DB 11921 CTGACCTTAGCAGCG 11935

RESULT 9
 US-08-658-136-2
 Sequence 2, Application US/08658136
 Patent No. 6071717
 GENERAL INFORMATION:
 APPLICANT: KLINGER, KATHERINE W
 APPLICANT: LANDES, GREGORY M
 APPLICANT: BURN, TIMOTHY C
 APPLICANT: CONNORS, TIMOTHY D
 APPLICANT: DACKOWSKI, WILLIAM
 APPLICANT: GERMINO, GREGORY
 APPLICANT: QIAN, FENG
 TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
 NUMBER OF SEQUENCES: 58
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: GENZYME CORPORATION
 STREET: ONE MOUNTAIN ROAD
 CITY: FRAMINGHAM
 STATE: MASSACHUSETTS
 COUNTRY: USA
 ZIP: 01701
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/658,136
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: LASSEN, ELIZABETH
 REGISTRATION NUMBER: 31,845
 REFERENCE/DOCKET NUMBER: GEN4-17.8
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 508-872-8400
 TELEFAX: 508-872-5415
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 53526 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 US-08-658-136-2

Alignment Scores:
 Pred. No.: 3,21e+03 Length: 53526
 Score: 81.00 Matches: 37
 Percent Similarity: 37.93% Conservative: 18
 Best Local Similarity: 25.52% Mismatches: 64
 Query Match: 10.09% Indels: 26
 DB: 1 Gaps: 8

US-09-868-025-2 (1-143) x US-08-658-136-2 (1-53526)

QY 11 GlnSerGluThrHisAlaTrpCysTrpSerSer-----SerThrThrArgSerProSer 28
 DB 12519 GAGGAGACACACATGTTCTTTTCTGGAGCTCTGAGTGGCCACGCGGACCCCGCC 12578

QY 29 ArgHisHisLeuHisArgGluArgIleProCysLeuAlaLeuGlyValThrAlaIleCys 48
 DB 12579 AAGCACCTGGAAATCAAAACATCCCGCTGCTCTGGGCTGGCTGCACTCTGCTGCT 12638

QY 49 SerLeuValTrpIleHisValSerHis-----GlyGlyGlyIleSerVal 63
 DB 12639 GCCTCCAGCTGGCTGAGCGCGGCGGCTGCTGGGCGCACAGCGGGGGCGCCACAGTC 12698

QY 64 AlaLeuCysSerGlnCysLeuGlnThrAsnAlaLeuArgProArgProAspCysLeuThr 83
 DB 12699 TCCCTGCAGAGTGAGCGGAGCTGGAAATGCACTCAGCCCTTTCAGAACACCTCGC 12758

QY 84 AsnAsn-----GlyGlyCysTyrglyGluCysHisGlySerLeuGlyHisValAspArg 101
Db 12759 TCCTCATGGCTGGAGCTGT-----CCTTGGCTAGGGGCC-----AGG 12797
QY 102 PheProGlnHis-----SerAsnGluTrpAsn---SerGlyMetAspSer 115
Db 12798 GTCCCGAGGACCTGGTGGCAGGAGAGAGCTACATCTGGGGCTGAGGGCGGCTGGGTCT 12857
QY 116 CysLysProLeuArgGlyGluPheLeuGlyValLeuThrProHisProLysMetGluPhe 135
Db 12858 TTCTCCCTCAGCTCCGAG-----GCCCGACCTGGGCCAGCGCTGGCATTC 12905
QY 136 AlaAlaIleArgala 140
Db 12906 CTGACCTTAGCAGCG 12920
RESULT 10
US-08-658-136-1
; Sequence 1, Application US/08658136
; Patent No. 6071717
; GENERAL INFORMATION:
; APPLICANT: KLINGER, KATHERINE W
; APPLICANT: LANDES, GREGORY M
; APPLICANT: BURN, TIMOTHY C
; APPLICANT: CONNORS, TIMOTHY D
; APPLICANT: DACKOWSKI, WILLIAM
; APPLICANT: GERMINO, GREGORY
; APPLICANT: QIAN, FENG
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: ONE MOUNTAIN ROAD
; CITY: FRAMINGHAM
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/658.136
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: LASSEN, ELIZABETH
; REGISTRATION NUMBER: 31,845
; REFERENCE/DOCKET NUMBER: GEN4-17.8
; TELEPHONE: 508-872-8400
; TELEFAX: 508-872-5415
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 53577 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-658-136-1
Alignment Scores:
Pred. No.: 3 21e+03 Length: 53577
Score: 81.00 Matches: 37
Percent Similarity: 37.93% Conservative: 18
Best Local Similarity: 25.52% Mismatches: 64
Query Match: 10.09% Indels: 26
DB: 3 Gaps: 8
US-09-868-025-2 (1-143) x US-08-658-136-1 (1-53577)

Db 12518 GAGGGAGACACACATGTTCTTTCTGGAGCTCTGCAGTGGCCACCGCAGCCGCC 12577
QY 29 ArgHisHisLeuHisArgGluArgIleProCysLeuAlaLeuGlyValThrAlaIleCys 48
Db 12578 AGCACCTCGGAATGAACATCCCGCTGCTGTCTGGGCTGGCTGCTGCTGCT 12637
QY 49 SerLeuValTrpIleHisValSerHis-----GlyGlyGlyIleSerVal 63
Db 12638 GCCTCCAGCTGCTGAGCCCGGCGACGCTCTCGGGCCACAGCGGGGCCACAGTC 12697
QY 64 AlaLeuCysSerGlnCysLeuGlnThrAsnAlaLeuArgProArgProAspCysLeuThr 83
Db 12698 TCCTTCAGAGTGGCGCAGCTGGAAATGACAGCTCAGCCCTTTCCAGAACACCTGCG 12757
QY 84 AsnAsn-----GlyGlyCysTyrglyGluCysHisGlySerLeuGlyHisValAspArg 101
Db 12758 TCCTCATGGCTGGAGCTGT-----CCTTGGCTAGGGGCC-----AGG 12796
QY 102 PheProGlnHis-----SerAsnGluTrpAsn---SerGlyMetAspSer 115
Db 12797 GTCCCGAGGACCTGGTGGCAGGAGAGAGCTACATCTGGGGCTGAGGGCGGCTGGGTCT 12856
QY 116 CysLysProLeuArgGlyGluPheLeuGlyValLeuThrProHisProLysMetGluPhe 135
Db 12857 TTCTCCCTCAGCTCCCGAG-----GCCCGACCTGGGCCAGCGCTGGCATTC 12904
QY 136 AlaAlaIleArgala 140
Db 12905 CTGACCTTAGCAGCG 12919
RESULT 11
US-09-371-008-1
; Sequence 1, Application US/09371008
; Patent No. 6127174
; GENERAL INFORMATION:
; APPLICANT: TONOUCHI, Naoto
; APPLICANT: SUGIYAMA, Masakazu
; APPLICANT: YOKOZAKI, Kenzo
; TITLE OF INVENTION: Plasmid Derived from Gluconobacter Bacteria
; TITLE OF INVENTION: and a Vector
; FILE REFERENCE: OP873
; CURRENT APPLICATION NUMBER: US/09/371,008
; CURRENT FILING DATE: 1999-08-10
; EARLIER APPLICATION NUMBER: JP 10-227227437
; EARLIER FILING DATE: 1998-08-11
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; TYPE: DNA
; ORGANISM: Gluconobacter oxydans
US-09-371-008-1
Alignment Scores:
Pred. No.: 185 Length: 5648
Score: 80.00 Matches: 30
Percent Similarity: 40.19% Conservative: 13
Best Local Similarity: 28.04% Mismatches: 28
Query Match: 9.96% Indels: 36
DB: 3 Gaps: 7
US-09-868-025-2 (1-143) x US-09-371-008-1 (1-5648)
QY 19 TrpSerSerThrThrArgSerProSerArgHisHisLeuHisArgGluArgIlePro 38
Db 1294 TGGACCATCAACACAGCGGCGCATCAAGATGGGGGTAGCCGCTCCCGC-----GCT 1350
QY 39 CysLeuAlaLeuGlyValThrAlaIleCysSerLeuValTrpIle----- 53
Db 1351 GCGCTCGCCCTTGCATGACGGGTCCGTGCAAAACCCCTTTGGGTATGCTGCTCAAGGGCCA 1410
QY 54 -----HisValSerHisGlyGlyIleSerValAlaLeuCysSer 67


```

DB      1358 TCCGGG-----ACGGCGTGGC-GCGGAATTC 1333

RESULT 13
US-08-319-866-1/c
; Sequence 1, Application US/08319866
; Patent No. 5929223
; GENERAL INFORMATION:
; APPLICANT: Tully, Timothy P.
; APPLICANT: Ylin, Jerry C.
; APPLICANT: Reguluski, Michael
; TITLE OF INVENTION: CLONING AND CHARACTERIZATION OF GENES
; TITLE OF INVENTION: ASSOCIATED WITH LONG-TERM MEMORY
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/319,866
; FILING DATE: 7-OCT-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: CSHL94-03
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1083 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "cDNA and PCR analysis"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1080
US-08-319-866-1

Alignment Scores:
Pred. No.: 24.4 Length: 1083
Score: 79.00 Matches: 28
Percent Similarity: 38.30% Conservative: 8
Best Local Similarity: 29.79% Mismatches: 38
Query Match: 9.84% Indels: 20
DB: Gaps: 4

US-09-868-025-2 (1-143) x US-08-319-866-1 (1-1083)
QY 17 TTPCyATPpSerSerThrThrArgSerProSerArgHisHis-----31
DB 665 TGGTGTGGGAATCGGTGTCGCACAGATTCGTGGCTGTCTCCGGGTGCTGGGTTC 606
QY 32 -----LeuHisArgGluAlaProCysLeuAlaLeuGlyValThrAla 46
DB 605 GGTTTGGGGCTTGATCTTACATCGGAAGGTTGGAGGGATTTTGTTCACGACTTGCAC 546
QY 47 IleCysSerLeuValTrpIleHisValSerHisGlyGly-----59
DB 545 TTCCACGGGGCTGGTGTGGATGACCTTCAGTTCCGGCGCTTGGCCACGTAGCACCTTC 486

```

Qy 60 GlyIleSerValAlaLeuCysSerGlnCysLeuGlnThrAsnAlaLeuArgProArgPro 79
Db 485 TGCATCGCTGGCGGCGCCAGCCCTGT-----TCTCTCTCTGGTTCAGCTGCT 432
Qy 80 AspCysLeuThrAsnAsnGly---GlyCysTyrGlyGluCys 92
Db 431 CTCGTATGACTCCCGAGGATGCGCTGTATCAGGACTGC 390

RESULT 14
US-09-124-758-5/c
; Sequence 5, Application US/09124758
; Patent No. 6146849
; GENERAL INFORMATION:
; APPLICANT: Pierce, J. M.
; APPLICANT: Moreman, Kelley W.
; TITLE OF INVENTION: Lectins and Coding Sequences
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/124,758
; FILING DATE: 04-JUN-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/048,507
; FILING DATE: 04-JUN-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 40-97
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1139 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 34..1011
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 34..87
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 88..1008
US-09-124-758-5

Alignment Scores:
Pred. No.: 26.1 Length: 1139
Score: 79.00 Matches: 29
Percent Similarity: 33.08% Conservative: 15
Best Local Similarity: 21.80% Mismatches: 49
Query Match: 9.84% Indels: 40
DB: 3 Gaps: 6

US-09-868-025-2 (1-143) x US-09-124-758-5 (1-1139)

Qy 21 SerSerThrThrArgSerProSerArgHisHisLeuHisArgGluArgIle----- 37
Db 739 AGTAATAAGATGCGCTCTTCTTAGCATACCAAGATCAGACACAGGATGCTGGCGG 680
Qy 38 -----ProCysLeuAlaLeuGlyValThrAlaIleCysSer 49
Db 679 CATTTGTCATTCCAAACATTCCCTGATCTGTATTTCATCTGGGTATTTCTGTAGATGCCAA 620
Qy 50 LeuValTrpIleHisValSerHisGlyGly-----GlyIleSerValAlaLeu 65
Db 619 ACAGATTATGTCTCCAGTCTCTGGAGGAAGCCAGTGTGTGGCGGTACCTCAGCAGGCGCG 560
Qy 66 Cys-----SerGlnCys---LeuGlnThrAsnAlaLeuArgProArg 78
Db 559 TGTTCCTCCAATGCTGCATGGGGACTTGTGTGGGCACATCCAGATGCCCGGCTCTTGG 500
Qy 79 ProAspCysLeuThrAsnAsnGlyCys----- 88
Db 499 CCTGGATGCTGATAGCCAGGCTTCTTGTAGTCATCGCTCGTGGCGGCTCTGCAGATC 440
Qy 89 -----TyrGlyGluCysHisGlySerLeuGlyHisValAspArgPhePro 103
Db 439 CAAAGGTGTTGTAGTTGGGCCAGTTCCTCCATCCCTCTGGGTAGTCTGCTTTGTTGGCCT 380
Qy 104 GlnHisSerAsnGluTrpAsnSerGlyMetAspSerCys 116
Db 379 GCT-----GACTGGACCAAGCGATCACCCACCGTGC 350

RESULT 15
US-09-709-677-5/c
; Sequence 5, Application US/09709677
; Patent No. 6524820
; GENERAL INFORMATION:
; APPLICANT: Pierce, J. M.
; APPLICANT: Moreman, Kelley W.
; Lee, Jin-Kyu
; TITLE OF INVENTION: Lectins and Coding Sequences
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/709,677
; FILING DATE: 09-No. 6524820-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/048,507 and US 09/124,758
; FILING DATE: 04-JUN-1997 and 04-JUN-1998, respectively.
; ATTORNEY/AGENT INFORMATION:
; NAME: Yoo-Warren, Heeja
; REGISTRATION NUMBER: 45,495
; REFERENCE/DOCKET NUMBER: 40-97a
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1139 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO

```

FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 34..1011
FEATURE:
/ NAME/KEY: sig_peptide
/ LOCATION: 34..87
FEATURE:
/ NAME/KEY: mat_peptide
/ LOCATION: 88..1008
/ SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-709-677-5

Alignment Scores:
Pred. No.: 26.1 Length: 1139
Score: 79.00 Matches: 29
Percent Similarity: 33.08% Conservative: 15
Best Local Similarity: 21.80% Mismatches: 49
Query Match: 9.84% Indels: 40
DB: 4 Gaps: 6

US-09-868-025-2 (1-143) x US-09-709-677-5 (1-1139)
Qy 21 SerSerThrThrArgSerProSerArgHisLeuHisArgGluArgile----- 37
Db 739 AGTAATAGATGCAAGTCTTCTAGCATCACCAAGTCATAGACCACAGGTATGGCTGGGC 680
Qy 38 -----ProCysLeuAlaLeuGlyValThrAlaIleCysSer 49
Db 679 CATGTGTCATTCCAACTTCCTGATCTGTTTCACTGGGTATTTCTGGTAGATGCCAA 620
Qy 50 LeuValTrpIleHisValSerHisGlyGly-----GlyIleSerValAlaLeu 65
Db 619 ACAGATTATGTCAGTCTCTGGAGGAGAGCCAGTGTGGTGGTACCTCAGCAGGCGC 560
Qy 66 Cys-----SerGlnCys---LeuGlnThrAsnAlaLeuArgProArg 78
Db 559 TGTTCCTCCAAATGCTGATGGGGGACTTGTGGGCACATGCCAGATGCCAGGTCCTTGG 500
Qy 79 ProAspCysLeuThrAsnAsnGlyGlyCys----- 88
Db 499 CCTGGATGTCGTAGTACCAGAGGTCTTGTAGTCATCGCTGGCGCGCCTCTGCAGATC 440
Qy 89 -----TyrGlyGluCysHisGlySerLeuGlyHisValAspArgPhePro 103
Db 439 CAAAGGTGTGTAGTGGCCCGCAGTTGCCATCCCGCTCTGGGTAGTCTGCTTTGTCCTT 380
Qy 104 GlnHisSerAsnGluTrpAsnSerGlyMetAspSerCys 116
Db 379 GCT-----GACTGGACCGAGGATCACCACCGTGC 350

```

Search completed: November 14, 2003, 14:05:14
Job time : 1432 secs

and is derived by analysis of the total score distribution.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_pzn model

Run on: November 14, 2003, 11:33:58 : Search time 218 Seconds
(without alignments)
1770.732 Million cell updates/sec

Title: US-09-868-025-2
Perfect score: 803
Sequence: 1 MVDTNFPISSESETHAWCMS.....LGVLTTPHKKMEFAAIRAGKV 143

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Zgapop 6.0, Zgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+pzn.model -DEV=xlh
-O=/cgn2_1/USPTO_spool/US09868025/runat_13112003_130405_17445/app_query.fasta_1.327
-DB=Geneseq_190003 -QFMT=fastp -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCLIGN=200 -THR SCORE=500 -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09868025 -CGN 1 1 0 @runat_13112003_130405_17445 -NCPUS=6 -ICPU=3
-NO MWAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOC=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq_190003.*
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	800	99.6	673	22	AHH19867
C 2	132	16.4	888	24	ABN98736
C 3	132	16.4	1247	21	AAC54660
C 4	132	16.4	1252	21	AAC54662
C 5	122	15.2	1253	21	AAC53587
C 6	113	14.1	1255	21	AAC54330
C 7	104	13.0	5350	24	ABK91137
C 8	99	12.3	1035	24	ABZ14447
C 9	99	12.3	1346	21	AAC54760
C 10	99	12.3	1349	21	AAC55078
C 11	91	11.3	1000	21	AAAE5508
C 12	90	11.2	18246	23	AAS35555
C 13	89.5	11.1	1092	18	AAT69760
C 14	89.5	11.1	1092	22	AAH27800
C 15	89.5	11.1	1483	20	AAZ52973
C 16	88.5	11.0	933	24	ABZ11909
C 17	88	11.0	583	22	AAK92045
C 18	88	11.0	583	22	AAK93708
C 19	88	11.0	3282	23	ABL13656
C 20	87	10.8	2627	22	AAK94378
C 21	87	10.8	4514	21	AAAS8867
C 22	86	10.7	438	25	ABX74750
C 23	86	10.7	1823	24	ABA05456
C 24	86	10.7	1838	22	AAK91317
C 25	86	10.7	1838	22	AAS32144
C 26	86	10.7	1838	24	ABN90499
C 27	86	10.7	2213	22	AAS46199
C 28	86	10.7	2213	25	ACA57957
C 29	86	10.7	2213	25	ABX98427
C 30	86	10.7	2213	25	ABX98929
C 31	86	10.7	2213	25	ACA05974
C 32	86	10.7	2213	25	ABX98018
C 33	86	10.7	2213	25	ABX78802
C 34	86	10.7	2213	25	ABX75815
C 35	86	10.7	2213	25	ABX77020
C 36	86	10.7	2213	25	ABX16860
C 37	86	10.7	2329	21	AAK75989
C 38	86	10.7	2348	25	ACC46288
C 39	86	10.7	2746	22	AAI59777
C 40	86	10.7	2845	22	AAH18012
C 41	85.5	10.6	13715	22	AAK83689
C 42	85.5	10.6	13715	23	ABK42411
C 43	85	10.6	5106	24	ABL60583
C 44	85	10.6	5437	24	ABL60382
C 45	84.5	10.5	9839	22	AAI99052

ALIGNMENTS

RESULT 1
AAH19867/c
ID AAH19867 standard; DNA; 673 BP.
XX
AC AAH19867;
XX
DT 03-AUG-2001 (first entry)
XX
DE Rice AGT-SAL 11 polynucleotide sequence SEQ ID NO:1.
XX
KW Rice AGT-SAL 11; salt tolerance; plant; cytosolic; antiHIV;
KW proteinase inhibitor; cancer; human immunodeficiency virus;
KW HIV infection; animal disorder; food processing; enzyme industry;
KW biological preservative; ds.
XX
OS Oryza sativa.
XX

PN WO200130990-A2.
 XX 03-MAY-2001.
 XX 11-OCT-2000; 2000WO-IN00099.
 XX 13-OCT-1999; 99IN-0000997.
 XX (AVES-) AVESTHAGEN GRAINE TECHNOLOGIES PVT LTD.
 XX Patell VM, Antony CM, Chandran D, Madurappa A;
 XX WPI; 2001-308632/32.
 XX P-PSDB; AAB5128.
 XX Nucleotide sequence encoding an AGT-SAL 11 polypeptide similar to
 PT Bowman Birk II type proteinase inhibitors is useful to confer salt
 PT resistance to plants -
 XX Claim 1; Page 17; 22pp; English.
 XX The present sequence encodes the rice AGT-SAL 11 protein. The AGT-SAL 11
 CC protein can be used to confer salt tolerance to plants and other
 CC organisms. The AGT-SAL 11 gene was isolated from salt-stressed rice.
 CC Also described are: (i) a transgenic plant comprising a recombinant
 CC expression cassette comprising a plant promoter operably linked to N1;
 CC and (ii) conferring salt tolerance on a plant, comprising introducing
 CC the above expression cassette. The AGT-SAL 11 protein is a proteinase
 CC inhibitor. The proteinase inhibitor may be used to confer stress
 CC tolerance to many plants and organisms including cotton, maize, rice,
 CC soybean, sugar beet, wheat, fruit, vegetables and vines, particularly
 CC biotic bacterial, fungal and pest stressors. Proteinase inhibitors are
 CC also useful in the treatment of cancer, human immunodeficiency virus
 CC (HIV) infection and other animal disorders. The gene may also be useful
 CC in food processing and enzyme industries as an inhibitor of proteinase
 CC activity and as a biological preservative.
 XX Sequence 673 BP; 163 A; 145 C; 170 G; 195 T; 0 other;

Alignment Scores:
 Pred. No.: 4,56e-66 Length: 673
 Score: 800.00 Matches: 142
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 99.30% Mismatches: 0
 Query Match: 99.63% Indels: 0
 DB: 22 Gaps: 0

US-09-868-025-2 (1-143) x AAH19867 (1-673)
 QY 1 MetValAspThrAnPheProIleSerGluInSerGluThrHisAlaTrpCysTrpSer 20
 Db 431 CTGTGATACAAATTTTCATTCAGTGCATCAGAGCAATCAGAAACATGTTGTGTGAGC 372
 QY 21 SerSerThrThrArgSerProSerArgHisLeuHisArgGluArgIleProCysLeu 40
 Db 371 AGCAGCACCAACAGAGCCGTCGCCCATCATCTTCATCGCAGCGGATCCCATGCGTT 312
 QY 41 AlaLeuGlyValThrAlaIleCysSerLeuValTrpIleHisValSerHisGlyGly 60
 Db 311 GCCCTCGGTGTCAGTCAATTTGAGTCTGTGTGGATATGTCACCCACGGTGTGGA 252
 QY 61 IleSerValAlaLeuCysSerGlnCysLeuGlnThrAsnAlaLeuArgProArgProAsp 80
 Db 251 ATCAGCGTGCCTTTTCAGGCGAGTGCCTTCAGACGAATGCGTCCGCCCTCGACCGAT 192
 QY 81 CysLeuThrAnAnGlyGlyCysTyGlyGlyCysHisGlySerLeuGlyHisValAsp 100
 Db 191 TGTCTGACCAACATGGAGGCGTGTATGTGTGAATGCCATGCGATGTTGGCATGTGAC 132
 QY 101 ArgPheProGlnHisSerAsnGluTrpAsnSerGlyMetAspSerCysLysProLeuArg 120
 Db 131 AGATTCTCAGCATTCATGATGATGAGACAGCGGCATGACAGCTGCAACACCATGAGG 72

QY 121 GlyGluPheLeuGlyValLeuThrProHisProLysMetGluPheAlaAlaIleArgAla 140
 Db 71 GGAGAAATTTCTGTGTGTCTACGCCACACCCCAAGATGGAGTTTGTGCGCATCCGACA 12
 QY 141 GlyLysVal 143
 Db 11 GGCAAGGTA 3
 RESULT 2
 ABN98736
 ID ABN98736 standard; DNA; 888 BP.
 XX AC ABN98736;
 XX DT 01-AUG-2002 (first entry)
 XX DE Arabidopsis thaliana expressed polynucleotide SEQ ID NO 504.
 XX KW Arabidopsis thaliana; plant; insecticide; fungicide; transgenic; stress;
 XX KW disease; crop; thale cress; tolerance factor; insect; pathogen;
 XX KW nutrition; ds.
 XX OS Arabidopsis thaliana.
 XX PN US2002023281-A1.
 XX PD 21-FEB-2002.
 XX PF 26-JAN-2001; 2001US-0770445.
 XX PR 27-JAN-2000; 2000US-178472P.
 XX (GRL/) GORLACH J.
 XX (ANY/) AN Y.
 XX (HML/) HAMILTON C M.
 XX (PRC/) PRICE J L.
 XX (RAIN/) RAINES T M.
 XX (YUY/) YU Y.
 XX (RAME/) RAMEAKA J G.
 XX (PAGE/) PAGE A.
 XX (MATH/) MATHEN A V.
 XX (LEDF/) LEDFORD B L.
 XX (WORS/) WORSNER J P.
 XX (HAAS/) HAAS W D.
 XX (GARC/) GARCIA C A.
 XX (KRIC/) KRICKER M.
 XX (SLAT/) SLATER T.
 XX (DAVI/) DAVIS K R.
 XX (ALLE/) ALLEN K.
 XX (HOFF/) HOFFMAN N.
 XX (HURB/) HURBAN P.
 XX Goriach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y;
 PI Rameaka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;
 PI Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;
 PI Hurban P;
 XX WPI; 2002-400781/43.
 XX New Arabidopsis thaliana nucleic acid for identifying homologous genes,
 PT producing compositions that modulate the expression or function of its
 PT encoded protein, and mapping functional regions of protein -
 XX Claim 1; SEQ ID NO 504; 49pp + Sequence Listing; English.
 XX The invention relates to an Arabidopsis thaliana nucleic acid (I)
 CC comprising a sequence capable of hybridising under stringent conditions
 CC to a sequence selected from any one of 999 sequences (ABN98233-ABN99231),
 CC given in the specification or its fragment. A polypeptide (II) encoded by
 CC (I), a transgenic plant (III) comprising an exogenous nucleic acid or a
 CC genetically modified cell (IV) comprising an exogenous nucleic acid, is
 CC useful for screening a candidate agent for its biological effect. (I) is
 CC useful in identifying homologous or related genes, in producing

compositions that modulate the expression or function of its encoded
protein, mapping functional regions of the protein and in studying
associated physiological pathways. (I) is also useful for the genetic
manipulation of cells, particularly plant cells. (II) is also useful in
screening assays of various plant strains to determine the strains that
are best capable of withstanding a particular disease or environmental
stress. (II) and (III) are useful for screening of biologically active
agents, e.g. fungicides, insecticides, etc., for elucidating biochemical
pathways. The screened agents are useful in improved methods of treating
crops to prevent or treat disease. (II) are also useful in screening
programs to identify agents that mimic or enhance the action of tolerance
factors. Such agents are useful in improved methods of treating crops to
enhance their tolerance to environmental stress. (I) is also useful
for enhancing or inhibiting production of a biosynthetic product in a
plant. (III) is useful for identifying other mediators that may induce
expression of proteins of interest, for establishing the extent to which
any specific insect and/or pathogen is responsible for damage to a
particular plant, for identifying other mediators that enhance or induce
tolerance to environmental stress, for identifying factors involved in
biosynthetic pathways of nutritional, commercial, or medicinal value and
for identifying productions of nutritional, commercial or medicinal
value. (IV) is useful in the study of genetic function and regulation,
CC for alteration of the cellular metabolism and for screening compounds
CC that may affect the biological function of the gene or gene products.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.htm?docID=99990977045.

XX SQ Sequence 888 BP; 257 A; 213 C; 171 G; 247 T; 0 other;

Alignment Scores:
Pred. No.: 0.00331 Length: 888
Score: 132.00 Matches: 30
Percent Similarity: 58.90% Conservative: 13
Best Local Similarity: 41.10% Mismatches: 29
Query Match: 16.44% Indels: 1
DB: 24 Gaps: 0

US-09-868-025-2 (1-143) x ABN98736 (1-888)

QY 67 SerGluCysLeuGlnThrAsnAlaLeuArgProArgProAspCysLeuThrAsnAsnGly 86
Db 307 AGCTACTGCTTTGAGACGGTAAGCGTCTTGACCCCAATCAGCTGCCAAGATCGA 366
QY 87 GlyCysTy-GlyGluCysHisGlySerLeuGlyHisValAspArgPheProGlnHisSer 106
Db 367 AGCTCCAATGGTCAGACCCATCGCTGTTTGGGCATCAGAACAGATTTCTTAAGCATGC 426
QY 107 AsnGluTrpAsnSerGlyMetAspSerCysLysProLeuArgGlyGluPheLeuGlyVal 126
Db 427 TATAAAGGCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 486
QY 127 LeuThrProHis-ProLysMetGluPheAlaAlaIle 138
Db 487 TTCTCTCCATACCTAGAGGCACATTTCAGCCATT 523

RESULT 3
AAC54660/c
ID AAC54660 standard; DNA; 1247 BP.

XX AC AAC54660;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 78636.
XX KW Hybridisation assay; genetic mapping; gene expression control;
XX KW protein identification; signal transduction pathway;
XX KW metabolic pathway; promoter; termination sequence; ss.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.

XX 06-SEP-2000. 2000EP-0301439.
PD 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0130891.
XX 30-APR-1999; 99US-0131419.
XX 30-APR-1999; 99US-0132048.
XX 04-MAY-1999; 99US-0132407.
XX 05-MAY-1999; 99US-0132484.
XX 06-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132563.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 18-MAY-1999; 99US-0134370.
XX 19-MAY-1999; 99US-0134768.
XX 20-MAY-1999; 99US-0134941.
XX 21-MAY-1999; 99US-0135124.
XX 24-MAY-1999; 99US-0135253.
XX 25-MAY-1999; 99US-0135829.
XX 28-MAY-1999; 99US-0136021.
XX 01-JUN-1999; 99US-0136782.
XX 03-JUN-1999; 99US-0137528.
XX 07-JUN-1999; 99US-0137502.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 17-JUN-1999; 99US-0139453.
XX 18-JUN-1999; 99US-0139492.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.
XX 18-JUN-1999; 99US-0139463.
XX 18-JUN-1999; 99US-0139750.
XX 21-JUN-1999; 99US-0139763.
XX 22-JUN-1999; 99US-0139817.
XX 23-JUN-1999; 99US-0139809.
XX 23-JUN-1999; 99US-0140353.
XX 24-JUN-1999; 99US-0140354.
XX 28-JUN-1999; 99US-0140695.
XX 29-JUN-1999; 99US-0140823.
XX 30-JUN-1999; 99US-0140991.
XX 01-JUL-1999; 99US-0141287.
XX 01-JUL-1999; 99US-0141842.
XX 01-JUL-1999; 99US-0142154.

```

PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0143202.
PR 12-JUL-1999; 99US-0143277.
PR 13-JUL-1999; 99US-0143642.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149830.
PR 26-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150684.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-SEP-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.

PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155459.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Alignment Scores:
Pred. No.: 0.00506
Score: 132.00
Percent Similarity: 58.90%
Best Local Similarity: 41.10%
Query Match: 16.44%
DB: 21

US-09-868-025-2 (1-143) x AAC54660 (1-1247)

Qy 67 SerGlnCysLeuGlnThrAsnAlaLeuArgProArgProAspCysLeuThrAsnAsnGly 86
Db 922 AGCTACTGCTTTGAGACGGTAACGCTCAGCTCTTGACCCAATCACCTGTCCCAAGATCGA 863
Qy 87 GlyCysTyrGlyGluCysHisGlySerLeuGlyHisValAspArgPheProGlnHisSer 106
Db 862 AGCTCCAATGGTCAAGCCATGGCTGTTTGGGCATCAGACAGATTTCCTAAGCATGTC 803
Qy 107 AsnGluTrpAsnSerGlyMetAspSerCysLeuArgGlyGluPheLeuGlyVal 126
Db 802 TATAAAGGCACAGCAGCATGGACGCCAAAACCATGACGGCGAAAACCTCTTTGTGTG 743
Qy 127 LeuThrProHis-ProLysMetGluPheAlaLalle 138
Db 742 TTCTCTCCATACCTAGAGGCACATTTCAGCCATT 706

RESULT 4
AAC54662/c
ID AAC54662 standard; DNA; 1252 BP.
XX AC AAC54662;
XX AC AAC54662;
Dt 18-OCT-2000 (first entry)

```

Arabidopsis thaliana DNA fragment SEQ ID NO: 78642.

Hybridisation assay; Genetic mapping; gene expression control;
protein identification; signal transduction pathway;
metabolic pathway; promote; termination sequence; ss.

Arabidopsis thaliana.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

25-FEB-1999; 99US-0121825.
05-MAR-1999; 99US-0123180.
09-MAR-1999; 99US-0123548.
23-MAR-1999; 99US-0125788.
25-MAR-1999; 99US-0126264.
29-MAR-1999; 99US-0126785.
01-APR-1999; 99US-0127462.
06-APR-1999; 99US-0128234.
08-APR-1999; 99US-0128714.
16-APR-1999; 99US-0129845.
19-APR-1999; 99US-0130077.
21-APR-1999; 99US-0130449.
23-APR-1999; 99US-0130510.
28-APR-1999; 99US-0130691.
28-APR-1999; 99US-0131449.
30-APR-1999; 99US-0132048.
30-APR-1999; 99US-0132407.
04-MAY-1999; 99US-0132484.
05-MAY-1999; 99US-0132485.
06-MAY-1999; 99US-0132486.
07-MAY-1999; 99US-0132487.
11-MAY-1999; 99US-0132863.
14-MAY-1999; 99US-0134256.
14-MAY-1999; 99US-0134218.
14-MAY-1999; 99US-0134219.
14-MAY-1999; 99US-0134221.
18-MAY-1999; 99US-0134370.
18-MAY-1999; 99US-0134378.
19-MAY-1999; 99US-0134941.
20-MAY-1999; 99US-0135124.
21-MAY-1999; 99US-0135353.
24-MAY-1999; 99US-0135629.
25-MAY-1999; 99US-0136021.
27-MAY-1999; 99US-0136392.
28-MAY-1999; 99US-0136782.
01-JUN-1999; 99US-0139119.
01-JUN-1999; 99US-0137222.
03-JUN-1999; 99US-0137528.
04-JUN-1999; 99US-0137502.
07-JUN-1999; 99US-0137724.
08-JUN-1999; 99US-0138094.
10-JUN-1999; 99US-0138540.
10-JUN-1999; 99US-0138847.
14-JUN-1999; 99US-0139119.
16-JUN-1999; 99US-0139452.
16-JUN-1999; 99US-0139453.
17-JUN-1999; 99US-0139492.
18-JUN-1999; 99US-0139454.
18-JUN-1999; 99US-0139455.
18-JUN-1999; 99US-0139456.
18-JUN-1999; 99US-0139457.
18-JUN-1999; 99US-0139458.
18-JUN-1999; 99US-0139459.
18-JUN-1999; 99US-0139460.
18-JUN-1999; 99US-0139461.
18-JUN-1999; 99US-0139462.
18-JUN-1999; 99US-0139463.
18-JUN-1999; 99US-0139750.
18-JUN-1999; 99US-0139751.

PR	27-AUG-1999;	99US-0151080.	Db	747 TTCTCCATACACCTAGAGGCACATTTCGAGCCATT 711	
PR	30-AUG-1999;	99US-0151333.	RESULT 5		
PR	31-AUG-1999;	99US-0151438.	AACS3587/c		
PR	01-SEP-1999;	99US-0151930.	ID AAC53587 standard; DNA; 1253 BP.		
PR	10-SEP-1999;	99US-0152363.	XX		
PR	10-SEP-1999;	99US-0153070.	AC AAC53587;		
PR	13-SEP-1999;	99US-0153758.	XX		
PR	15-SEP-1999;	99US-0154018.	DT 18-OCT-2000 (first entry)		
PR	16-SEP-1999;	99US-0154039.	XX		
PR	20-SEP-1999;	99US-0154779.	XX		
PR	22-SEP-1999;	99US-0155133.	XX		
PR	23-SEP-1999;	99US-0155486.	XX		
PR	24-SEP-1999;	99US-0155659.	XX		
PR	28-SEP-1999;	99US-0156458.	KW Hybridisation assay; genetic mapping; gene expression control;		
PR	29-SEP-1999;	99US-0156596.	KW protein identification; signal transduction pathway;		
PR	04-OCT-1999;	99US-0157117.	KW metabolic pathway; promoter; termination sequence; ss.		
PR	05-OCT-1999;	99US-0157753.	XX		
PR	06-OCT-1999;	99US-0157865.	OS Arabidopsis thaliana.		
PR	07-OCT-1999;	99US-0158029.	XX		
PR	08-OCT-1999;	99US-0158232.	PN BP1033405-A2.		
PR	12-OCT-1999;	99US-0158369.	XX		
PR	13-OCT-1999;	99US-0159293.	PD 06-SEP-2000.		
PR	13-OCT-1999;	99US-0159294.	XX		
PR	13-OCT-1999;	99US-0159295.	PF 25-FEB-2000; 2000EP-0301439.		
PR	14-OCT-1999;	99US-0159329.	XX		
PR	14-OCT-1999;	99US-0159330.	XX		
PR	14-OCT-1999;	99US-0159331.	PR 25-FEB-1999;		
PR	14-OCT-1999;	99US-0159337.	PR 05-MAR-1999;		
PR	14-OCT-1999;	99US-0159636.	PR 09-MAR-1999;		
PR	18-OCT-1999;	99US-0159584.	PR 23-MAR-1999;		
PR	21-OCT-1999;	99US-0160767.	PR 25-MAR-1999;		
PR	21-OCT-1999;	99US-0160768.	PR 29-MAR-1999;		
PR	21-OCT-1999;	99US-0160770.	PR 01-APR-1999;		
PR	21-OCT-1999;	99US-0160814.	PR 06-APR-1999;		
PR	21-OCT-1999;	99US-0160815.	PR 08-APR-1999;		
PR	21-OCT-1999;	99US-0160880.	PR 16-APR-1999;		
PR	22-OCT-1999;	99US-0160881.	PR 19-APR-1999;		
PR	22-OCT-1999;	99US-0160889.	PR 21-APR-1999;		
PR	23-OCT-1999;	99US-0161404.	PR 23-APR-1999;		
PR	25-OCT-1999;	99US-0161405.	PR 28-APR-1999;		
PR	25-OCT-1999;	99US-0161406.	PR 30-APR-1999;		
PR	26-OCT-1999;	99US-0161359.	PR 04-MAY-1999;		
PR	26-OCT-1999;	99US-0161360.	PR 05-MAY-1999;		
PR	26-OCT-1999;	99US-0161361.	PR 06-MAY-1999;		
PR	28-OCT-1999;	99US-0161320.	PR 07-MAY-1999;		
PR	28-OCT-1999;	99US-0161392.	PR 11-MAY-1999;		
PR	28-OCT-1999;	99US-0161393.	PR 14-MAY-1999;		
PR	29-OCT-1999;	99US-0162142.	PR 14-MAY-1999;		
Alignment Scores:			PR 14-MAY-1999;		
Pred. No.:	0.00509	Length: 1252	PR 14-MAY-1999;		
Score:	132.00	Matches: 30	PR 18-MAY-1999;		
Percent Similarity:	58.90%	Conservative: 13	PR 19-MAY-1999;		
Best Local Similarity:	41.10%	Mismatches: 29	PR 20-MAY-1999;		
Query Match:	16.44%	Indels: 1	PR 21-MAY-1999;		
DB:	21	Gaps: 0	PR 24-MAY-1999;		
US-09-868-025-2 (1-143) x AAC54662 (1-1252)			PR 25-MAY-1999;		
QY	67	serGlnCysLeuGlnThrAenAlaLeuArgProArgProAspCysLeuThrAenAsnGly 86	PR 27-MAY-1999;		
Db	927	AGCTACTGCCTTGAGACGGTAACGCTCAGCTCTTGACCCCAATCAGCTGCCAAGATCGA 868	PR 28-MAY-1999;		
QY	87	GlyCysTyrglyGlyCysHisGlySerLeuGlyHisValAspArgPheProGlnHisSer 106	PR 03-JUN-1999;		
Db	867	AGCTCCAATGCTCAAGCCATGGCTGTTTGGGCATCAGAACAGATTTCTTAAGCATTCG 808	PR 04-JUN-1999;		
QY	107	AsnGlnTrpAenSerGlyMetAspSerCysLeuArgGlyGluPheLeuGlyVal 126	PR 07-JUN-1999;		
Db	807	TATAAAGGCACAGCATGACAGCAGCAAAACCATGACAGGCGAAGAACTCTTTGTGTG 748	PR 08-JUN-1999;		
QY	127	LeuThrProHis-ProLysMetGluPheAlaAlaIle 138	PR 10-JUN-1999;		

PR 18-JUN-1999;	99US-0139456.	PR 18-AUG-1999;	99US-0149426.
PR 18-JUN-1999;	99US-0139457.	PR 20-AUG-1999;	99US-0149722.
PR 18-JUN-1999;	99US-0139458.	PR 20-AUG-1999;	99US-0149723.
PR 18-JUN-1999;	99US-0139459.	PR 20-AUG-1999;	99US-0149929.
PR 18-JUN-1999;	99US-0139460.	PR 23-AUG-1999;	99US-0149902.
PR 18-JUN-1999;	99US-0139461.	PR 23-AUG-1999;	99US-0149930.
PR 18-JUN-1999;	99US-0139462.	PR 25-AUG-1999;	99US-0150566.
PR 18-JUN-1999;	99US-0139463.	PR 26-AUG-1999;	99US-0150884.
PR 18-JUN-1999;	99US-0139750.	PR 27-AUG-1999;	99US-0151065.
PR 18-JUN-1999;	99US-0139753.	PR 27-AUG-1999;	99US-0151066.
PR 18-JUN-1999;	99US-0139773.	PR 27-AUG-1999;	99US-0151080.
PR 21-JUN-1999;	99US-0139817.	PR 30-AUG-1999;	99US-0151303.
PR 21-JUN-1999;	99US-0139899.	PR 31-AUG-1999;	99US-0151438.
PR 23-JUN-1999;	99US-0140353.	PR 01-SEP-1999;	99US-0151930.
PR 23-JUN-1999;	99US-0140354.	PR 07-SEP-1999;	99US-0152363.
PR 23-JUN-1999;	99US-0140695.	PR 10-SEP-1999;	99US-0153070.
PR 24-JUN-1999;	99US-0140823.	PR 13-SEP-1999;	99US-0153758.
PR 28-JUN-1999;	99US-0140821.	PR 15-SEP-1999;	99US-0154018.
PR 29-JUN-1999;	99US-0140991.	PR 16-SEP-1999;	99US-0154039.
PR 30-JUN-1999;	99US-0141287.	PR 20-SEP-1999;	99US-0154779.
PR 01-JUL-1999;	99US-0141842.	PR 22-SEP-1999;	99US-0155119.
PR 01-JUL-1999;	99US-0142154.	PR 23-SEP-1999;	99US-0155486.
PR 02-JUL-1999;	99US-0142055.	PR 24-SEP-1999;	99US-0155659.
PR 06-JUL-1999;	99US-0142390.	PR 28-SEP-1999;	99US-0156458.
PR 08-JUL-1999;	99US-0142803.	PR 29-SEP-1999;	99US-0156596.
PR 09-JUL-1999;	99US-0142920.	PR 04-OCT-1999;	99US-0157117.
PR 12-JUL-1999;	99US-0143542.	PR 05-OCT-1999;	99US-0157753.
PR 13-JUL-1999;	99US-0143622.	PR 06-OCT-1999;	99US-0157865.
PR 14-JUL-1999;	99US-0144003.	PR 07-OCT-1999;	99US-0158029.
PR 15-JUL-1999;	99US-0144085.	PR 08-OCT-1999;	99US-0158212.
PR 16-JUL-1999;	99US-0144086.	PR 12-OCT-1999;	99US-0158389.
PR 19-JUL-1999;	99US-0144325.	PR 13-OCT-1999;	99US-0159293.
PR 19-JUL-1999;	99US-0144332.	PR 13-OCT-1999;	99US-0159294.
PR 19-JUL-1999;	99US-0144332.	PR 13-OCT-1999;	99US-0159295.
PR 19-JUL-1999;	99US-0144333.	PR 14-OCT-1999;	99US-0159339.
PR 19-JUL-1999;	99US-0144333.	PR 14-OCT-1999;	99US-0159330.
PR 19-JUL-1999;	99US-0144334.	PR 14-OCT-1999;	99US-0159331.
PR 19-JUL-1999;	99US-0144335.	PR 14-OCT-1999;	99US-0159637.
PR 20-JUL-1999;	99US-0144352.	PR 14-OCT-1999;	99US-0159638.
PR 20-JUL-1999;	99US-0144632.	PR 14-OCT-1999;	99US-0159638.
PR 20-JUL-1999;	99US-0144884.	PR 18-OCT-1999;	99US-0159584.
PR 21-JUL-1999;	99US-0144814.	PR 18-OCT-1999;	99US-0159584.
PR 21-JUL-1999;	99US-0145086.	PR 21-OCT-1999;	99US-0160741.
PR 21-JUL-1999;	99US-0145088.	PR 21-OCT-1999;	99US-0160767.
PR 21-JUL-1999;	99US-0145085.	PR 21-OCT-1999;	99US-0160768.
PR 22-JUL-1999;	99US-0145087.	PR 21-OCT-1999;	99US-0160770.
PR 22-JUL-1999;	99US-0145089.	PR 21-OCT-1999;	99US-0160814.
PR 22-JUL-1999;	99US-0145089.	PR 21-OCT-1999;	99US-0160815.
PR 22-JUL-1999;	99US-0145192.	PR 22-OCT-1999;	99US-0160980.
PR 23-JUL-1999;	99US-0145145.	PR 22-OCT-1999;	99US-0160981.
PR 23-JUL-1999;	99US-0145218.	PR 22-OCT-1999;	99US-0160989.
PR 23-JUL-1999;	99US-0145224.	PR 25-OCT-1999;	99US-0161404.
PR 26-JUL-1999;	99US-0145276.	PR 25-OCT-1999;	99US-0161405.
PR 27-JUL-1999;	99US-0145913.	PR 25-OCT-1999;	99US-0161406.
PR 27-JUL-1999;	99US-0145918.	PR 26-OCT-1999;	99US-0161359.
PR 27-JUL-1999;	99US-0145919.	PR 26-OCT-1999;	99US-0161360.
PR 28-JUL-1999;	99US-0145951.	PR 26-OCT-1999;	99US-0161361.
PR 02-AUG-1999;	99US-0145951.	PR 26-OCT-1999;	99US-0161361.
PR 02-AUG-1999;	99US-0145986.	PR 28-OCT-1999;	99US-0161920.
PR 02-AUG-1999;	99US-0145988.	PR 28-OCT-1999;	99US-0161920.
PR 02-AUG-1999;	99US-0146388.	PR 28-OCT-1999;	99US-0161992.
PR 03-AUG-1999;	99US-0146389.	PR 28-OCT-1999;	99US-0161992.
PR 04-AUG-1999;	99US-0147038.	PR 29-OCT-1999;	99US-0161993.
PR 04-AUG-1999;	99US-0147204.	PR 29-OCT-1999;	99US-0162142.
PR 05-AUG-1999;	99US-0147302.		
PR 05-AUG-1999;	99US-0147192.		
PR 05-AUG-1999;	99US-0147260.		
PR 06-AUG-1999;	99US-0147303.		
PR 06-AUG-1999;	99US-0147416.		
PR 09-AUG-1999;	99US-0147493.		
PR 09-AUG-1999;	99US-0147935.		
PR 10-AUG-1999;	99US-0148171.		
PR 11-AUG-1999;	99US-0148319.		
PR 12-AUG-1999;	99US-0148341.		
PR 12-AUG-1999;	99US-0148565.		
PR 13-AUG-1999;	99US-0148684.		
PR 16-AUG-1999;	99US-0149368.		
PR 17-AUG-1999;	99US-0149175.		

Alignment Scores:
 Pred. No.: 0.0442
 Score: 122.00
 Percent Similarity: 56.1%
 Best Local Similarity: 35.7%
 Query Match: 15.1%
 DB: 21

US-09-868-025-2 (1-143) x AAC53587 (1-1253)

Qy 67 SerGlnCysLeuGlnThrAsnAlaLeuArgProAcgProAcgCysLeuThrAsnAsnGly 86
 Db 928 AGCTACTGCTTTGACGACGGTACGCTCAGCTTTGACCCCAATCACCCTGCTCCCAAGATCGA 869

QY	87	GlyCysTyrGlyGluCysHisGlySerLeuGlyHisValAspArgPheProGlnHisSer	106
DB	868	AGCTCCAATGTCAAAGCCATGGCTGTTTGGGCATCAGCACAGATTTCCIAAGCATTGC	809
QY	107	AsnGluTrpAsnSerGlyMetAspSerCysGlyPheProLeuArgGlyGluPheLeuGlyVal	126
DB	808	TATAAAGGCACACACGATGGACTGGCCAAACCATGAAGCGCAAACTTCTTTGTGTG	749
QY	127	LeuThrProHis-ProLysMetGluPheAlaAlaIle	138
DB	748	TTCTCTCATACACCTAGAGGCACATTTCGAGGCATT	712
RESULT 6			
AAC54330/c			
ID	AAC54330 standard; DNA; 1255 BP.		
XX	AAC54330:		
DT	18-OCT-2000 (first entry)		
XX	Arabidopsis thaliana DNA fragment SEQ ID NO: 77495.		
DE	Hybridisation assay; genetic mapping; gene expression control;		
XX	Protein identification; signal transduction pathway;		
KW	metabolic pathway; promoter; termination sequence; ss.		
XX	Arabidopsis thaliana.		
OS	EP1033405-A2.		
PN	06-SEP-2000.		
XX	25-FEB-2000; 2000EP-0301439.		
PD	25-FEB-1999; 99US-0121825.		
XX	05-MAR-1999; 99US-0123180.		
PR	09-MAR-1999; 99US-0123548.		
PR	23-MAR-1999; 99US-0125788.		
PR	25-MAR-1999; 99US-0126264.		
PR	29-MAR-1999; 99US-0126785.		
PR	01-APR-1999; 99US-0127462.		
PR	06-APR-1999; 99US-0128234.		
PR	08-APR-1999; 99US-0128714.		
PR	15-APR-1999; 99US-0129845.		
PR	19-APR-1999; 99US-0130077.		
PR	21-APR-1999; 99US-0130449.		
PR	23-APR-1999; 99US-0130510.		
PR	23-APR-1999; 99US-0130691.		
PR	28-APR-1999; 99US-0131449.		
PR	30-APR-1999; 99US-0132048.		
PR	30-APR-1999; 99US-0132407.		
PR	04-MAY-1999; 99US-0132484.		
PR	05-MAY-1999; 99US-0132485.		
PR	06-MAY-1999; 99US-0132486.		
PR	05-MAY-1999; 99US-0132487.		
PR	07-MAY-1999; 99US-0132863.		
PR	11-MAY-1999; 99US-0134256.		
PR	14-MAY-1999; 99US-0134218.		
PR	14-MAY-1999; 99US-0134219.		
PR	14-MAY-1999; 99US-0134221.		
PR	14-MAY-1999; 99US-0134370.		
PR	18-MAY-1999; 99US-0134768.		
PR	19-MAY-1999; 99US-0134941.		
PR	20-MAY-1999; 99US-0135124.		
PR	21-MAY-1999; 99US-0135353.		
PR	24-MAY-1999; 99US-0135629.		
PR	25-MAY-1999; 99US-0136021.		
PR	27-MAY-1999; 99US-0136592.		
PR	28-MAY-1999; 99US-0136782.		
PR	01-JUN-1999; 99US-0137222.		
PR	03-JUN-1999; 99US-0137528.		
PR	04-JUN-1999; 99US-0137502.		

Percent Similarity: 57.53% Conservatives: 12
Best Local Similarity: 41.10% Mismatches: 30
Query Match: 14.07% Indels: 2
DB: 21 Gaps: 0

US-09-868-025-2 (1-143) x AAC54330 (1-1255)

Qy 67 SerGlnCysLeuGlnThrAsnAlaLeuArgProArgProAspCysLeuThrAsnGly 86
Db 930 AGCTACTGCTTTGAGACGGTAAAGCTTACCTTTGACCAT -CACCTGTCACAGATCA 872
Qy 87 GlyCysTyrglyGluCysHisGlySerLeuGlyHisValAspArgPheProGlnHisser 106
Db 871 AGCTCCAATGGTCAAAGCCATGCTGTTTGGGATCAGACAGATTTCCTAAGCATTC 812
Qy 107 AsnGluTrpAsnSerGlyMetAspSerCysLysProLeuArgGlyGluPheLeuGlyVal 126
Db 811 TATAAAGGCACAGACATGGACACCAAAACCATGCGAGCGAAACCTTCTTTGTGTG 752
Qy 127 LeuThrProHis-ProGlyMetGluPheAlaAlaile 138
Db 751 TTCTCTCCATACCTAGAGGCACATTTCAGCCATT 715

RESULT 7
ABK93137
ID ABK93137 standard; cDNA; 5350 BP.
XX AC ABK93137;
XX 22-AUG-2002 (first entry)
XX Human prostate specific nucleic acid #47.
XX Prostate specific nucleic acid; PSNA; ss; prostate cancer; cytostatic;
XX vaccine.
XX Homo sapiens.
XX WO200239431-A2.
XX 16-MAY-2002.
XX 06-NOV-2001; 2001WO-US47175.
XX 06-NOV-2000; 2000US-246039P.
XX (DIAD-) DIADEXUS INC.
XX Sun Y, Recipon H, Chen S, Liu C;
XX WPI; 2002-471567/50.
XX New polypeptide useful for diagnosing and monitoring the presence and
XX metastases of prostate cancer in a patient, and for use as a component
XX in databases for search analysis as well as in sequence analysis
XX algorithms -
XX Claim1; Page 176-178; 257pp; English.
XX The invention relates to an isolated polypeptide (I) comprising a
XX sequence with 60 % identity to 107 sequences (prostate specific
XX proteins, PSP) (SI). Also included are the nucleic acids encoding the
XX PSPs (prostate specific nucleic acids, PSNA), vectors, host cells
XX antibodies and the use of the PSP or PSNA as a vaccine.
XX PSP and PSNA are useful for diagnosing and monitoring the presence and
XX metastases of prostate cancer in a patient. PSNA is useful for
XX determining the presence of a prostate specific nucleic acid (PSNA) in a
XX sample. An antibody that binds PSP is useful for determining the presence
XX of prostate specific protein in a sample, and for treating a patient with
XX prostate cancer, which induces an immune response against the prostate
XX cancer cell expressing the nucleic acid or polypeptide and a kit is
XX useful for detecting a risk of cancer or presence of cancer in a patient.
XX PSNA is useful as a hybridisation probe to detect, characterise and

99US-0147416.
99US-0147493.
99US-0147935.
99US-0148171.
99US-0148319.
99US-0148341.
99US-0148565.
99US-0148584.
99US-0149368.
99US-0149175.
99US-0149426.
99US-0149722.
99US-0149723.
99US-0149929.
99US-0149802.
99US-0149920.
99US-0150566.
99US-0150884.
99US-0151065.
99US-0151066.
99US-01511080.
99US-0151103.
99US-0151438.
99US-0151930.
99US-0152363.
99US-0153070.
99US-0153758.
99US-0154018.
99US-0154019.
99US-0154779.
99US-0155139.
99US-0155496.
99US-0155559.
99US-0156458.
99US-0156596.
99US-0157117.
99US-0157753.
99US-0157865.
99US-0158029.
99US-0158232.
99US-0158369.
99US-0159293.
99US-0159294.
99US-0159295.
99US-0159329.
99US-0159330.
99US-0159331.
99US-0159637.
99US-0159638.
99US-0159584.
99US-0160741.
99US-0160767.
99US-0160768.
99US-0160770.
99US-0160814.
99US-0160815.
99US-0160980.
99US-0160981.
99US-0160989.
99US-0161404.
99US-0161405.
99US-0161406.
99US-0161359.
99US-0161360.
99US-0161361.
99US-0161920.
99US-0161922.
99US-0161993.
99US-0162142.

Alignment Scores:
Pred. No.: 0.31
Score: 113.00

Length:
Matches: 1255
30

XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0125785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128714.
PR 08-APR-1999; 99US-0128845.
PR 16-APR-1999; 99US-0130077.
PR 19-APR-1999; 99US-0130449.
PR 21-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134321.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135829.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138847.
PR 10-JUN-1999; 99US-0139119.
PR 14-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139839.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 24-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142134.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142330.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149802.
PR 23-AUG-1999; 99US-0149920.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151530.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154019.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.

PR 24-JUN-1999;	99US-0140695.	PR 07-SEP-1999;	99US-0152363.		
PR 28-JUN-1999;	99US-0140823.	PR 10-SEP-1999;	99US-0153070.		
PR 29-JUN-1999;	99US-0140991.	PR 13-SEP-1999;	99US-0153758.		
PR 30-JUN-1999;	99US-0141287.	PR 15-SEP-1999;	99US-0154018.		
PR 01-JUL-1999;	99US-0141842.	PR 16-SEP-1999;	99US-0154039.		
PR 01-JUL-1999;	99US-0142154.	PR 20-SEP-1999;	99US-0154779.		
PR 02-JUL-1999;	99US-0142055.	PR 22-SEP-1999;	99US-0155139.		
PR 06-JUL-1999;	99US-0142055.	PR 23-SEP-1999;	99US-0155486.		
PR 08-JUL-1999;	99US-0142803.	PR 24-SEP-1999;	99US-0155659.		
PR 09-JUL-1999;	99US-0142920.	PR 28-SEP-1999;	99US-0156458.		
PR 12-JUL-1999;	99US-0142977.	PR 29-SEP-1999;	99US-0156596.		
PR 13-JUL-1999;	99US-0143542.	PR 04-OCT-1999;	99US-0157117.		
PR 14-JUL-1999;	99US-0143624.	PR 05-OCT-1999;	99US-0157753.		
PR 15-JUL-1999;	99US-0144005.	PR 06-OCT-1999;	99US-0157865.		
PR 16-JUL-1999;	99US-0144085.	PR 07-OCT-1999;	99US-0158029.		
PR 16-JUL-1999;	99US-0144086.	PR 08-OCT-1999;	99US-0158232.		
PR 19-JUL-1999;	99US-0144325.	PR 12-OCT-1999;	99US-0158369.		
PR 19-JUL-1999;	99US-0144331.	PR 13-OCT-1999;	99US-0159293.		
PR 19-JUL-1999;	99US-0144332.	PR 13-OCT-1999;	99US-0159294.		
PR 19-JUL-1999;	99US-0144333.	PR 13-OCT-1999;	99US-0159295.		
PR 19-JUL-1999;	99US-0144334.	PR 14-OCT-1999;	99US-0159329.		
PR 19-JUL-1999;	99US-0144335.	PR 14-OCT-1999;	99US-0159330.		
PR 20-JUL-1999;	99US-0144332.	PR 14-OCT-1999;	99US-0159331.		
PR 20-JUL-1999;	99US-0144612.	PR 14-OCT-1999;	99US-0159637.		
PR 20-JUL-1999;	99US-0144884.	PR 18-OCT-1999;	99US-0159638.		
PR 21-JUL-1999;	99US-0144884.	PR 18-OCT-1999;	99US-0159584.		
PR 21-JUL-1999;	99US-0145086.	PR 21-OCT-1999;	99US-0160741.		
PR 21-JUL-1999;	99US-0145088.	PR 21-OCT-1999;	99US-0160767.		
PR 21-JUL-1999;	99US-0145085.	PR 21-OCT-1999;	99US-0160768.		
PR 22-JUL-1999;	99US-0145085.	PR 21-OCT-1999;	99US-0160770.		
PR 22-JUL-1999;	99US-0145089.	PR 21-OCT-1999;	99US-0160814.		
PR 22-JUL-1999;	99US-0145192.	PR 21-OCT-1999;	99US-0160815.		
PR 23-JUL-1999;	99US-0145145.	PR 22-OCT-1999;	99US-0160980.		
PR 23-JUL-1999;	99US-0145218.	PR 22-OCT-1999;	99US-0160981.		
PR 23-JUL-1999;	99US-0145224.	PR 22-OCT-1999;	99US-0160989.		
PR 26-JUL-1999;	99US-0145276.	PR 25-OCT-1999;	99US-0161404.		
PR 27-JUL-1999;	99US-0145913.	PR 25-OCT-1999;	99US-0161405.		
PR 27-JUL-1999;	99US-0145918.	PR 25-OCT-1999;	99US-0161406.		
PR 27-JUL-1999;	99US-0145919.	PR 26-OCT-1999;	99US-0161359.		
PR 28-JUL-1999;	99US-0145951.	PR 26-OCT-1999;	99US-0161360.		
PR 02-AUG-1999;	99US-0146386.	PR 26-OCT-1999;	99US-0161361.		
PR 02-AUG-1999;	99US-0146388.	PR 28-OCT-1999;	99US-0161320.		
PR 02-AUG-1999;	99US-0146389.	PR 28-OCT-1999;	99US-0161392.		
PR 03-AUG-1999;	99US-0147038.	PR 28-OCT-1999;	99US-0161393.		
PR 04-AUG-1999;	99US-0147204.	PR 29-OCT-1999;	99US-0162142.		
PR 05-AUG-1999;	99US-0147302.				
PR 05-AUG-1999;	99US-0147312.				
PR 06-AUG-1999;	99US-0147260.				
PR 06-AUG-1999;	99US-0147303.				
PR 06-AUG-1999;	99US-0147416.				
PR 09-AUG-1999;	99US-0147493.				
PR 09-AUG-1999;	99US-0147915.				
PR 10-AUG-1999;	99US-0148171.				
PR 11-AUG-1999;	99US-0148319.				
PR 12-AUG-1999;	99US-0148341.				
PR 13-AUG-1999;	99US-0148565.				
PR 13-AUG-1999;	99US-0148684.				
PR 16-AUG-1999;	99US-0149368.				
PR 17-AUG-1999;	99US-0149175.				
PR 18-AUG-1999;	99US-0149426.				
PR 20-AUG-1999;	99US-0149722.				
PR 20-AUG-1999;	99US-0149723.				
PR 20-AUG-1999;	99US-0149929.				
PR 23-AUG-1999;	99US-0149902.				
PR 23-AUG-1999;	99US-0149930.				
PR 25-AUG-1999;	99US-0150566.				
PR 26-AUG-1999;	99US-0150884.				
PR 27-AUG-1999;	99US-0151065.				
PR 27-AUG-1999;	99US-0151066.				
PR 27-AUG-1999;	99US-0151080.				
PR 30-AUG-1999;	99US-0151303.				
PR 31-AUG-1999;	99US-0151438.				
PR 01-SEP-1999;	99US-0151930.				
		Alignment Scores:			
		Pred. No.:	7	Length:	1349
		Score:	99.00	Matches:	25
		Percent Similarity:	49.25%	Conservative:	8
		Best Local Similarity:	37.31%	Mismatches:	34
		Query Match:	12.33%	Indels:	0
		DB:	21	Gaps:	0
US-09-868-025-2 (1-143) x AAC35078 (1-1349)					
QY	65	LeuCysSerGlnCysLeuGlnThrAsnAlaLeuArgProArgProAspCysLeuThrAsn	84		
DB	1015	CTTCTCAGCTACAGACTTGAGTCTCGCTCTCTGCTCTTGACCAATCACCCTGTCCTCA	956		
QY	85	AsnGlyCysArgGlyGlyCysHisGlySerLeuGlyHisValAspArgPheProGln	104		
DB	955	AACGTATGCTGCTATGCTGATGATACCACTGCTTTTAGCATCAACACTGCTTTCTCTCAG	896		
QY	105	HisSerAsnGlnTrpAsnSerGlyMetAspSerCysValProLeuArgGlyGluPheLeu	124		
DB	895	TATTCCTATGAATGAACCGCTCGTGAAGAGCTATAACCAAGATGCCGMAAATCTTTC	836		
QY	125	GlyValLeuThrProHisPro	131		
DB	835	AGTGTGCTCTCTCCAGACCCC	815		
RESULT 11					

AAA65508/c
ID AAA65508 standard; DNA; 1000 BP.
XX
AC AAA65508;
XX
XX
DT 10-NOV-2000 (first entry)
XX
DE Porcine BAC-PIGF2-1 contig 99.
XX
KW Porcine; pig; wild boar; quantitative trait locus; QTL; chromosome 2;
KW mapping; 2p1.7; select breeding; genotype; phenotype; muscle mass;
KW fat deposition; IGF2; insulin-like growth factor 2; ds.
XX
XX
OS Sus scrofa.
XX
XX WO200036143-A2.
XX
XX 22-JUN-2000.
XX
XX 16-DEC-1999; 99WO-EP10209.
XX
XX 16-DEC-1998; 98EP-0204231.
XX
XX (UYLI-) UNIV LIEGE.
XX (MELI-) MELICA HB.
XX (SEGH-) SEGHERSGENTEC NV.
XX
XX Andersson L, Georges M, Spincemalle G;
XX WPI; 2000-431612/37.
XX
XX Selecting a domestic animal for having desired genotypic properties
XX comprises testing the animal for the presence of a parentally imprinted
XX quantitative trait locus which is related to muscle mass and/or fat
XX deposition -
XX
XX Example 3; Fig 6; 107pp; English.
XX
XX The present invention describes a method (M1) for selecting a domestic
XX animal for having desired genotypic properties. The method comprises
XX testing the animal for the presence of a parentally imprinted
XX quantitative trait locus (QTL). The pig QTL is located at chromosome 2,
XX mapping at around position 2p1.7. Also described are: (1) an isolated
XX and/or recombinant nucleic acid (N1) comprising a parentally imprinted
XX QTL or its functional fragment; (2) an isolated and/or recombinant
XX nucleic acid (N2) comprising a phenotypically parentally imprinted QTL
XX derived from at least one chromosome or its functional fragment; (3) an
XX animal such as pig selected for having desired genotypic or potential
XX phenotypic properties; (4) a transgenic animal comprising N1 or N2; and
XX (5) sperm or an embryo derived from the animal of (3) or (4). N1 or its
XX fragment is useful for selecting an animal destined for slaughter or a
XX breeding animal having desired genotypic or potential phenotypic
XX properties. The properties are related to muscle mass and/or fat
XX deposition. The sperm or an embryo are useful in breeding animals
XX destined for slaughter. AAA65418 to AAA65524 represent contigs 1 to 10
XX and 19 to 115 which were isolated from porcine BAC-PIGF2-1 which
XX contains the INS and IGF2 (insulin-like growth factor) genes. These
XX sequences were used in an example from the present invention for
XX generating a reference sequence of IGF2 and flanking loci in the pig.
XX
XX Sequence 1000 BP; 180 A; 223 C; 337 G; 260 T; 0 other;
SQ

Alignment Scores:
Pred. No.: 27.1 Length: 1000
Score: 91.00 Matches: 40
Percent Similarity: 33.96% Conservative: 14
Best Local Similarity: 25.16% Mismatches: 57
Query Match: 11.33% Indels: 48
DB: 21 Gaps: 6

US-09-868-025-2 (1-143) x AAA65508 (1-1000)

Qy 12 SerGluThrHisAlaTrpCysTrpSerSerThrThrArgSerProSerArgHis--- 30

DB 775 GCAATACACATGACACAGCATTCTACTACACACACGTTAGACACATGATGACACACA 716
QY 31 -----HisLeuHisArgGluArgileProCysLeuAlaLeuGlyValThrAlaileCys 48
DB 715 CATGCACACACACAGGCAC-----AATGCACACAGC 683
QY 49 SerLeuValTrpIleHisValSerHisGlyGlyIleSer---ValAlaLeuCysSer 67
DB 682 ACACCTCAGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 623
QY 68 GlncysLeuGlnThrAsnAlaLeuArgProArgProArgProArgProArgProArgProArg 87
DB 622 CACACGCTGCTGTGGGGTAACCTTTTCCACGCTCCAGGCTGCAAAAGTAGAAGCAGCTCCG 563
QY 88 CysTyrglyGlyCysHisGlySer-----LeuGly 97
DB 562 TGTGAGGGTGGTTAGTGGTCTGGGCCCGCAGTCCTCCCGGCACTCTTTTGGCCAGC 503
QY 98 HisValAspArgpHeProGlnHis-----SerAsn-Gl 108
DB 502 TCTCTGCTCTCAGTTTCCCCATCACCACCCACCTCCACCTCCCAAGGTAACCTGGAGACCCA 443
QY 108 utrpLenSer-----GlyMetAs 114
DB 442 GTGGGACATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 383
QY 114 pSerCysLysProLeuArgGly-GluPheLeuGlyValLeuThrProHisPro 131
DB 382 GCTCTGTGCCCCCAGGAGGGGCGACACCTTGGACCTTAGTAATCTCTGCCCGC 330

RESULT 12
AAS59555
ID AAS59555 standard; DNA; 18246 BP.
XX
XX AAS59555;
XX
XX 13-FEB-2002 (first entry)
XX
XX Propionibacterium acnes immunogenic protein encoding DNA #50.
XX
XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
XX inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
XX dermatological; osteopathic; neuroprotectant; ds.
XX
XX Propionibacterium acnes.
XX
XX WO200181581-A2.
XX
XX 01-NOV-2001.
XX
XX 20-APR-2001; 2001WO-US12865.
XX
XX 21-APR-2000; 2000US-199047P.
XX 02-JUN-2000; 2000US-208841P.
XX 07-JUL-2000; 2000US-216747P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiry YAM, Parsing DH, Mitcham JL, Wang SS, Bhatia A;
XX L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
XX WPI; 2001-616774/71.
XX
XX
XX Propionibacterium acnes polypeptides and nucleic acids useful for
XX vaccinating against and diagnosing infections, especially useful for
XX treating acne vulgaris -
XX
XX Claim 1; SEQ ID No 50; 1069pp; English.
XX
XX Sequences AAS59506-AAS59804 represent DNA molecules encoding
XX Propionibacterium acnes immunogenic polypeptides. The proteins and their

CC associated DNA sequences are used in the treatment, prevention and
 CC diagnosis of medical conditions caused by P. acnes. The disorders include
 CC SAPHO syndrome (synovitis, acne, pustulosis, hyperostosis and
 CC osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved
 CC in infections of bone, joints and the central nervous system, however it
 CC is particularly involved in the inflammatory lesions associated with acne
 CC vulgaris. A method for detecting the presence or absence of P. acnes in a
 CC patient comprises contacting a sample with a binding agent that binds to
 CC the proteins of the invention and determining the amount of bound protein
 CC in the sample. The polypeptides may be used as antigens in the production
 CC of antibodies specific for P. acnes proteins. These antibodies can be
 CC used to downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA). This sequence encodes the
 CC polypeptides shown in AAU52458-AAU52625 and AAU67539.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pat_sequences.

XX Sequence 18246 BP; 4200 A; 4737 C; 5096 G; 4213 T; 0 other;

Alignment Scores:
 Pred. No.: 1,27e+03 Length: 18246
 Score: 90.00 Matches: 40
 Percent Similarity: 39.39% Conservative: 12
 Best Local Similarity: 30.30% Mismatches: 51
 Query Match: 11.21% Indels: 29
 DB: 23 Gaps: 5

US-09-868-025-2 (1-143) x AAS59555 (1-18246)

QY 15 HisAlaTrpCysTrpSerSerThrThrArgSerProSerArg----- 29
 DB 16590 CACATCTGGATCGAAGAGGATCCGAAGCCGT---CCAGTAGGAATGATGCGGTCG 16646
 QY 30 -----HisHisLeuHisArgGluArgPro-----CysLeuAla 41
 DB 16647 GAACCGTATCAGTGTGATGTGAAGCCCTACCCITACGAGTCAAGCATCTTGTGTCGT 16706
 QY 42 LeuGlyValThrAlaIleCysSerLeuValTrpIleHisValSerHisGlyGlyLe 61
 DB 16707 CCAGAAGCCACTCTCATTTTGTCTTTCATTGGGAAGTGT-----GGGAC 16754
 QY 62 SerValAlaLeuCysSerGlnCys-LeuGlnThrAsnAlaLeuArgProArgProAspCy 81
 DB 16755 TCTGTCGGCAGGTGTACTGCTGTTTTCATTACTAATATCGAGCGTAGCAGACCTGACGT 16814
 QY 81 sLeuThrAsnAsnGlyGlyCysTrpGlyGlyCysHis----- 93
 DB 16815 GATGTCGATCGACTTGGCTGTAGTGGTCTACCCACCTTTTACGACACTCCAGCACT 16874
 QY 94 -GlySerLeuGly--HisValAspArgPheProGlnHisSerAsnGlnTrpAsnSerGly 112
 DB 16875 AGCCAGCTTGTGTGACATTGCCACAGACGAGGAGGAGCACTTCTCCTCAGCTCCGTAGT 16934
 QY 113 MetAspSerCysIleProLeuArgGlyGlu 122
 DB 16935 GTGCAGCAGTGCAGCACTGATGCGCAAG 16964

RESULT 13
 AAT69760
 ID AAT69760 standard; cDNA; 1092 BP.
 XX
 AC AAT69760;
 XX
 DT 09-SEP-1997 (first entry)
 XX
 DE Rat growth hormone secretagogue receptor type Ia cDNA.
 XX
 KW Growth hormone secretagogue receptor; GHSR;
 XX G protein coupled receptor; gs.

OS Rattus sp.
 XX Key Location/Qualifiers
 XX CDS 1..1092
 XX /tag= a
 XX /transl_except= pos: 190..190, aa:Asn
 XX /transl_except= pos: 364..366, aa:Ser

W09721730-A1.
 XX 19-JUN-1997.
 XX 10-DEC-1996; 96WO-US19445.
 XX 06-JUN-1996; 96US-0018962.
 XX 13-DEC-1995; 95US-0008582.
 XX (MERI) MERCK & CO INC.

XX Arena JP, Cully DF, Feighner SD, Howard AD, Liberator PA;
 XX Schaeffer JM, Van Der Ploeg L;
 XX WPI; 1997-332725/30.
 XX P-PSDB; AAM19220.
 XX Receptor of growth hormone receptor family - specifically
 XX secretagogue or secretagogue like receptor, useful to screen for
 XX specific binding agents for growth hormone deficiency treatment

XX Claim 26; Fig 24A-B; 70pp; English.

XX A DNA sequence (AAT69760) comprises the coding sequence for rat type
 XX Ia growth hormone secretagogue receptor (GHSR) (AAM19220). A
 XX full-length sequence, including an intron (see also AAT69759) was
 XX obt'd. by screening a rat pituitary library with pig GHSR clone 7-3
 XX (see also AAT69754) and by PCR amplification of the N-terminal
 XX fragment. Pig and human GHSR clones (see also AAT69754-58) were
 XX similarly identified. GHSRs can be used to screen for specific
 XX binding agents useful in the treatment of conditions related to
 XX growth hormone shortage.

XX Sequence 1092 BP; 187 A; 359 C; 297 G; 249 T; 0 other;

Alignment Scores:
 Pred. No.: 41.9 Length: 1092
 Score: 89.50 Matches: 38
 Percent Similarity: 35.21% Conservative: 12
 Best Local Similarity: 26.76% Mismatches: 50
 Query Match: 11.15% Indels: 42
 DB: 18 Gaps: 7

US-09-868-025-2 (1-143) x AAT69760 (1-1092)

QY 17 TrpCysTrpSerSerSerThrThr-ArgSerProSerArgHisLeuHisArgGluAr 36
 DB 542 TGGCGCTGGAGCAGCAAAACCGCAGATCCCGGGACACCAAGATGCCCGCCACCG 601
 QY 36 gileProCysLeuAlaLeuGlyValThrAlaIleCysSerLeuValTrpIleHisValse 56
 DB 602 AGTTCCTGT-----GGCTCTGGGTGCTCACCG 631
 QY 56 rHisGlyGlyGlyLe-----SerValAlaLeuCysSerGlnCys---Le 70
 DB 632 TCATGTGTGGTGTCTCCAGCGTCTTCTTCTACCGGTCTTCTGCTCTACTGTGCTCT 691
 QY 70 uGlnThrAsn-----Al 74
 DB 692 ACAGTCTCATCGGAGGAGCTAATGCCGAGACCGGAGATGCAGCGGTGGCGCTCGC 751
 QY 74 aLeuArgProArgProAspCysLeuThrAsnAsnGlyGlyCysTrpGlyGlyCysHisG 94
 DB 752 TCCGGGACCAAGACCAACAGCAGACAGATGCTTCTGTGTGTGTGTGTGTGTGTTC 811

QY 94 ySerLeu-----GlyHisValaspArgPheProGlnHisSerAsnG1 108
 Db 812 TCCTCTGCTGGTCCCTCCAGTGGGAGATACCTCTTTTC---CAAGTCCCTCGAGC 868
 QY 108 utpAsnSerGlyMetAspSerCysLysProLeuAspGlyGluPheLeuGlyValLeuTh 128
 Db 869 CTGGCTCTCTGGAGATCGCTCAGATCAGCCAGTACTGCAACCT-----GGTGTCTTTG 922
 QY 128 rPro 129
 Db 923 TCCT 926
 RESULT 14
 AAH27800
 ID AAH27800 standard; DNA; 1092 BP.
 AC AAH27800;
 DT 16-AUG-2001 (first entry)
 XX Rat growth hormone secretagogue receptor (GHSR) related protein DNA.
 DE Growth hormone secretagogue receptor; GHSR; GHSR ligand; neotropic;
 KW antiarthritic; cytostatic; antidiabetic; pituitary dwarfism;
 KW Turner's syndrome; chronic nephritis; chondrodystrophy; Down's syndrome;
 KW Silver syndrome; bone formation disorder; juvenile chronic arthritis;
 KW megacephalia; tumour; insulinoma; calcinoid; diabetes; rat; ds.
 XX
 OS Ratt's sp.
 XX
 XX Location/Qualifiers
 FT Key 1..1263
 FT CDS /tag= a
 FT /partial
 FT /product= "GHSR-related protein"
 FT /note= "Growth hormone secretagogue receptor-related
 protein, no stop codon given"
 XX
 XX WO200132705-A1.
 PD 10-MAY-2001.
 PF 31-OCT-2000; 2000WO-JP07635.
 PR 01-NOV-1999; 99JP-0311632.
 PR 17-DEC-1999; 99JP-0358723.
 XX (TAKE) TAKEDA CHEM IND LTD.
 XX
 XX Hinuma S, Kawamata Y, Fukusumi S, Fujii R;
 PI WPI; 2001-308741/32.
 DR P-PSDB; AAB97377.
 XX
 XX Growth-hormone secretagogue receptor ligand polypeptides and encoded
 PT DNAs, useful in diagnosis and screening drug candidates for treating
 PT e.g. pituitary dwarfism, Turner syndrome, chronic nephritis
 XX
 PS Disclosure: Page 100-101; 106pp; Japanese.
 XX
 XX This invention relates to a growth hormone secretagogue receptor (GHSR)
 CC ligand peptide. The invention contains bovine, human, rat and porcine
 CC GHSR gene and protein sequences, and also GHSR ligand sequences. The use
 CC of an antibody directed against GHSR ligand, GHSR promoters and
 CC inhibitors may lead to neotropic, antiarthritic, cytostatic, and
 CC antidiabetic activity. GHSR DNA, protein, promoters and inhibitors are
 CC useful for preventing or treating diseases caused by insufficient growth
 CC hormone, e.g. pituitary dwarfism, Turner syndrome, chronic nephritis,
 CC chondrodystrophy, adult pituitary failure, Down's syndrome, Silver
 CC syndrome, bone formation disorder, juvenile chronic arthritis syndrome or
 CC for diseases of megacephalia, thyroid stimulating hormone (TSH)-promoting
 CC tumours, non-secretory (non-functional) pituitary tumours, ectopic
 CC adrenocorticotrophic hormone (ACTH)-producing tumours, myeloid thymoma,
 CC

CC vasotropic inhibitory peptide (VIP)-producing tumours, glucagon-producing
 CC tumour, gastrin-producing tumour, insulinoma, calcinoid,
 CC insulin-dependent or non-dependent diabetes. The present sequence
 CC represents DNA encoding a rat GHSR-related protein.

SQ Sequence 1092 BP; 186 A; 358 C; 239 G; 249 T; 0 other;

Alignment Scores:
 Pred. No.: 41.9 Length: 1092
 Score: 89.50 Matches: 38
 Percent Similarity: 35.21% Conservative: 12
 Best Local Similarity: 26.76% Mismatches: 50
 Query Match: 11.15% Indels: 42
 Gaps: 7

US-09-868-025-2 (1-143) x AAH27800 (1-1092)

QY 17 TipCysTrpSerSerThrThr-ArgSerProSerArgHisHisLeuHisArgGluAr 36
 Db 542 TGGCGTGGAGCAGCAAAACGGCACAGATCCCGGAGACACCAAGATCCCGCCACCG 501
 QY 36 qileProCysLeuAlaLeuGlyValThrAlaIleCysSerLeuValTrpIleHisValse 56
 Db 602 AGTTCGGTGT-----GGCTCTGGGTGCTCACCG 631
 QY 56 rHisGlyGlyGlyLe-----SerValAlaLeuCysSerGlnCys---Le 70
 Db 632 TCATGGTGTGGGTGTCCAGCGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 591
 QY 70 uGlnThrAsp-----A1 74
 Db 592 ACAGTCTCATCGGAGGAGCACTATGCGGAGACGCGAGATGCGGCGGCGCTCGC 751
 QY 74 alaArgProArgProAspCysLeuThrAsnAnGlyGlyCysTyGlyGlyCysHisG1 94
 Db 752 TCCGGGACCAAGAACCAACAGACAGATGAAGATGCTTCTGCTGGTGGTGGTGGTTC 811
 QY 94 ySerLeu-----GlyHisValaspArgPheProGlnHisSerAsnG1 108
 Db 812 TCCTCTGCTGGTGGCTTCCAGCTGGGAGATACCTCTTTTC---CAAGTCTTCGAGC 868
 QY 108 utpAsnSerGlyMetAspSerCysLysProLeuAspGlyGluPheLeuGlyValLeuTh 128
 Db 869 CTGGCTCTCTGGAGATCGCTCAGATCAGCCAGTACTGCAACCT-----GGTGTCTTTG 922
 QY 128 rPro 129
 Db 923 TCCT 926
 RESULT 15
 AA252973/C
 ID AA252973 standard; cDNA; 1483 BP.
 XX AC AA252973;
 XX
 XX 14-MAR-2000 (first entry)
 DT Human prostate tumor cDNA library derived EST fragment #116.
 DE
 XX Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;
 XX treatment; ds.
 XX Homo sapiens.
 OS
 XX DB19820190-A1.
 PN
 XX 04-NOV-1999.
 PD
 XX 28-APR-1998; 98DE-1020190.
 PP
 XX 28-APR-1998; 98DE-1020190.
 PR
 XX (META-) METAGEN GES GENOMFORSCHUNG MBH.
 PA

```

XX Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl B;
PI WPI; 1999-621386/54.
XX DR P-PSDB; AAY74163, AAY74164, AAY74165.
XX PT New human nucleic acid sequences from pancreatic tumors, and related
XX PT proteins
XX PS Claim 2; Page 277; 502pp; German.
XX CC This invention describes novel polypeptides and their encoding nucleic
XX CC acids derived from human pancreatic tumor tissue which have cytostatic
XX CC activity. The sequences are also useful in producing pharmaceutical
XX CC compositions for treatment of pancreatic tumors. AA252858-253014
XX CC represent expressed sequence tag (EST) fragments derived from a human
XX CC pancreatic tumor cDNA library and which encode the proteins represented
XX CC in AAY73814-Y74252.
XX SQ Sequence 1483 BP; 309 A; 405 C; 420 G; 349 T; 0 other;

Alignment Scores:
Pred. No.: 61.4 Length: 1483
Score: 89.50 Matches: 37
Percent Similarity: 36.30% Conservative: 16
Best Local Similarity: 25.34% Mismatches: 59
Query Match: 11.15% Indels: 34
DB: Gaps: 8

US-09-868-025-2 (1-143) x AA252973 (1-1483)
QY 14 ThrHisAlaTrpCysTrpSerSerThrThrArgSerProSerArgHisHis----- 31
DB 1034 ACACCTGGCTGGACGAGCAGATCACGCCAGCCAGCTCCACGAGGACCCAGCGC 975
QY 32 ---LeuHisArgGluArgIleProCysLeuAlaLeuGlyValThrAlaIleCysSerLeu 50
DB 974 ACTACCCACAGCGTGCACATTCG-----GCAGGGCCCC 942
QY 51 ValTrpIleHisValSerHisGlyGlyIleSerValAlaLeuCysSerGlnCysLeu 70
DB 941 TCCTGG-----GTGCCCCACGGGAGGGCTCGGGGTC-----ATC 906
QY 71 GlnThrAsnAlaLeuArgPro-ArgProAspCysLeu-----Th 83
DB 905 TCCAGCACTGCCACAGGCGCGCCAGCCAGATTCTTGGGAACACCCGCTCGGAAC 846
QY 83 rAsnAsnGlyGlyCysTyrGlyGlu-----CysHisGlySerLeuGlyHisValAsp-- 100
DB 845 CGAAGGTCAGGGGTGTCCTTCCAGAGCGCGTGTCTCCAGGGTCCGAGGCGCGCGG 786
QY 101 ----ArgPheProGlnHisSerAsnGluTrpAsnSerGlyMetAspSerCys-----Ly 117
DB 785 CTGACACAGCCCCCGGGTGGAGGTCTCTGGCAGGCCAGCTTCCAGCATGCCGACGCA 726
QY 117 sProLeuArgGlyGluPheLeuGlyValLeuThrProHisProLysMetGluPheAlaAl 137
DB 725 GCCTAGGGAGCCCGGCTCTGTGGCATCTTCACTCTGAGCCTCAGAGCCCCCATCTGAA 666
QY 137 aIleArgAlaGlyLys 142
DB 665 CGGCAAGACCGGAGG 650

```

Search completed: November 14, 2003, 12:33:42
 Job time : 226 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 14, 2003, 11:35:29 ; Search time 2089 Seconds
(without alignments)
2800.420 Million cell updates/sec

Title: us-09-868-025-2

Perfect score: 803

Sequence: 1 MVDNPPISQSETHAWCMS.....LGLVTPHPKMEFAAIRAGKV 143

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US0868025/runat_13112003_130405_17456/app_query.fasta_1.327
-DR=GenEmb1 -QPM=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blowsum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORW=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US0868025@cgn 1 1 2372 @runat_13112003_130405_17456 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NGM_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPT=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmb1:
1: gb_ba:
2: gb_htg:
3: gb_in:
4: gb_om:
5: gb_ov:
6: gb_sat:
7: gb_ph:
8: gb_pi:
9: gb_pr:
10: gb_ro:
11: gb_sts:
12: gb_sy:
13: gb_un:
14: gb_vi:
15: em_ba:
16: em_fun:
17: em_hum:
18: em_in:
19: em_mu:
20: em_om:
21: em_or:
22: em_ov:
23: em_pat:
24: em_ph:
25: em_pi:
26: em_ro:
27: em_sts:
28: em_un:

29: em_vi:
30: em_htg_hum:
31: em_htg_inv:
32: em_htg_other:
33: em_htg_mus:
34: em_htg_pln:
35: em_htg_rtd:
36: em_htg_mam:
37: em_htg_vrt:
38: em_sy:
39: em_htgo_hum:
40: em_htgo_mus:
41: em_htgo_other:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	800	99.6	673	8	AF192975 Oryza sat
2	744	92.7	123673	8	AP002837 Oryza sat
3	631.5	78.6	80327	2	AP003955 Oryza sat
4	631.5	78.6	191580	8	AP003847 Oryza sat
5	610.5	78.0	1184	6	AX653397 Sequence
6	148.5	18.5	1104	6	AX654660 Sequence
7	148.5	18.5	95893	8	AB017123 Oryza sat
8	148.5	18.5	139999	8	AC018727 Oryza sat
9	132	16.4	1060	8	BT002755 Arabidops
10	132	16.4	1180	8	AY050908 Arabidops
11	132	16.4	85020	8	AB018113 Arabidops
12	128	15.9	148522	2	AC126015 Arabidops
13	122	15.2	1275	8	AY088538 Arabidops
14	106.5	13.3	174874	10	AC123054 Mus muscu
15	106.5	13.3	206926	10	AC124497 Mus muscu
16	105.5	13.1	167676	2	AC113153 Homo sapi
17	105.5	13.1	235357	9	AC103758 Homo sapi
18	105	13.1	206125	10	AL672219 Mouse DNA
19	100.5	12.5	674	1	UBQ408154 Unculture
20	100.5	12.5	674	1	UBQ408228 Unculture
21	100	12.5	673	1	UBQ408089 Unculture
22	100	12.5	172388	2	AC137258 Rattus no
23	100	12.5	193495	2	AC122100 Rattus no
24	100	12.5	252313	2	AC127110 Rattus no
25	100	12.5	262917	2	AC097239 Rattus no
26	100	12.5	291668	2	AC117849 Rattus no
27	99	12.3	1035	6	AX507557 Sequence
28	99	12.3	1035	6	AX651340 Sequence
29	99	12.3	1349	8	AY085639 Arabidops
30	99	12.3	6232	10	AB037890 Mus muscu
31	99	12.3	99688	8	ATFL3M23 Arabidops
32	99	12.3	197070	8	ATCHR1V62 Arabidops
33	98.5	12.3	208202	2	AC013446 Homo sapi
34	98.5	12.3	329718	2	AC114465 Homo sapi
35	97	12.1	92444	9	AX088650 Homo sapi
36	96.5	12.0	167479	9	AL445235 Human DNA
37	96.5	12.0	214946	9	AC136896 Homo sapi
38	95	11.8	111864	2	AC139687 Homo sapi
39	95	11.8	113764	10	AL928599 Mouse DNA
40	95	11.8	164113	2	AC068128 Homo sapi
41	95	11.8	169125	9	AC126605 Homo sapi
42	95	11.8	175310	9	AC135735 Homo sapi
43	95	11.8	193680	9	AC127482 Homo sapi
44	95	11.8	195966	2	AC102785 Mus muscu
45	94.5	11.8	1582	9	AF052032 Homo sapi

ALIGNMENTS

RESULT 1

AF192975/c
 LOCUS Oryza sativa unknown gene. 673 bp DNA linear PLN 09-NOV-1999
 DEFINITION Oryza sativa unknown gene.
 ACCESSION AF192975
 VERSION AF192975.1 GI:6289051
 KEYWORDS Oryza sativa (indica cultivar-group)
 SOURCE Oryza sativa (indica cultivar-group)
 ORGANISM Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
 REFERENCE 1 (bases 1 to 673)
 AUTHORS Patel, V.M., Mathai, C.A., Divya, C. and Ashok, M.
 TITLE Oryza sativa Variety IR64 (cDNA clone AGTSAL-11 from 7 days old seedling)
 JOURNAL unpublished
 REFERENCE 2 (bases 1 to 673)
 AUTHORS Patel, V.M., Mathai, C.A., Divya, C. and Ashok, M.
 TITLES Direct Submission
 JOURNAL Submitted (08-OCT-1999) Plant Genome Biology Department, Avesthagen Grain Technologies, P.Box 5091, Cubbon Park G.P.O., Bangalore, Karnataka 560001, India
 FEATURES
 source
 1..673
 /organism="Oryza sativa (indica cultivar-group)"
 /mol_type="genomic DNA"
 /cultivar="IR64"
 /db_xref="taxon:39946"
 /clone="AGTSAL-11"
 /tissue_type="7 day old seedling"
 /complement("1..>431")
 /product="unknown"
 /complement("1..431")
 /function="role in salinity stress response"
 /codon_start=1
 /product="unknown"
 /protein_id="AF06789.1"
 /db_xref="GI:6289052"
 /translation="MVDNFPISQSETHAWCSSTSPSRHLLHREIPICLALGV
 TAICSLVTHVHGSGISVALCSQCLTNALRPDPCLTNNGGICGCHSGLSGVDRF
 PQHSNEMNSGMDSCPLRGEFLGVLTPHPQEFPAIRAGV"
 BASE COUNT 163 a 145 c 170 g 195 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,33e-62 Length: 673
 Score: 800.00 Matches: 142
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 99.30% Mismatches: 0
 Query Match: 99.63% Indels: 0
 DB: Gaps: 0
 US-09-868-025-2 (1-143) x AF192975 (1-673)
 Qy 1 MetValaspThrAnPheProIleSerGluGlnSerGluThrHisAlaTrpCysTrpSer 20
 Db 431 CTGGTAGATACAAATTTCCATCATCAGTGCAGCAACACATGCTGGTGTGGAGC 372
 Qy 21 SerSerThrThrArgSerProSerArgHisHisLeuHisArgGluArgIleProCysLeu 40
 Db 371 AGCAGACCCACAGAGACCGCTCCGCCATCATCTTCATCGGAGCGGATCCCATGCTT 312
 Qy 41 AlaLeuGlyValThrAlaIleCysSerLeuValTrpIleHisValSerHisGlyGlyGly 60
 Db 311 GCCCTCGGTGCATCGCAATTTGAGCTTTGTTGGATACATGTCAGCCACCGTGTGGA 252
 Qy 61 IleSerValAlaLeuCysSerGlnCysLeuGlnThrAsnAlaLeuArgProArgProAsp 80
 Db 251 ATCAGCGTCGCTTTTGCAGCAGTGCCTTCAGACGGAATGCGTCCGCCCTCGACCCGAT 192
 Qy 81 CysLeuThrAnAnGlyGlyCysThrGlyCysHisGlySerLeuGlyHisValAsp 100
 Db 191 TGTCTGACCAACAATGAGGCGTGTATGGTGAATGCCATGGCATGCTTGGGCATGTGTGAC 132

Qy 101 AtgPheProGlnHisSerAsnGluTrpAsnSerGlyMetAspSerCysLysProLeuArg 120
 Db 131 AGATTTCCTCAGCATTCCTCAATGAATGACAGCGGCGCATGCACAGCTGCAACCACTGAG 72
 Qy 121 GlyCluPheLeuGlyValLeuThrProHisProLysMetGluPheAlaLeuArgAla 140
 Db 71 GGAGAAATTTCTGGTGTGCTCAGCCACACCCACAGATGGAGTTTCTCCATCCGAGCA 12
 Qy 141 GlyLysVal 143
 Db 11 GGCAAGGTA 3
 RESULT 2
 AP002837 123673 bp DNA linear PLN 26-OCT-2002
 LOCUS Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 6,
 DEFINITION BAC clone: OSJNBa0019P11.
 ACCESSION AP002837
 VERSION AP002837.2 GI:24413940
 KEYWORDS Oryza sativa (japonica cultivar-group)
 SOURCE Oryza sativa (japonica cultivar-group)
 ORGANISM Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
 REFERENCE 1
 AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.
 TITLES Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, BAC clone: OSJNBa0019P11
 JOURNAL Published Only in Database (2000)
 REFERENCE 2 (bases 1 to 123673)
 AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.
 TITLES Direct Submission
 JOURNAL Submitted (03-AUG-2000) Takuji Sasaki, National Institute of Agrobiological Sciences, Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
 (E-mail: tsasaki@nias.affrc.go.jp, URL: http://xgp.dna.affrc.go.jp/, Tel: 81-298-38-7441, Fax: 81-298-38-7468)
 On Oct 26, 2002 this sequence version replaced GI:9711842.
 Genes were predicted from the integrated results of the following:
 GENSCAN (http://ccr-081.mit.edu/GENSCAN.html), FGENESH (http://www.softberry.com/), GeneMark.hmm (http://opal.biology.gatech.edu/GeneMark/), GlimmerM (http://opal.tigr.org/tdb/glimmer/glmr.form.html), RiceHMM (http://xgp.dna.affrc.go.jp/RiceHMM/), SplicePredictor (http://bioinformatics.iastate.edu/cgi-bin/sp.cgi), BLASTN and BLASTX. The genomic sequence was searched against NCBI NonRedundant Protein database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at RGP. Protein homologues of the coding regions were searched against NCBI NonRedundant Protein database with BLASTP. ESTs represent the identified cDNA sequences using BLASTN with the corresponding DBJ accession no. and RGP clone ID.
 A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative', and '-like protein'. A gene without significant homology to any protein but with RGP homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted by two or more gene prediction programs is classified as a 'hypothetical' protein according to IRGSP standard. A gene predicted by a single gene prediction program is also classified as a probable 'hypothetical' protein and is included as a miscellaneous feature of the sequence.
 The orientation of the sequence is from -21M13 to M13rev of the BAC clone. This sequence of OSJNBa0019P11 clone has an overlap with P054LH01 (DBJ:AP001389) clone at the 5' end. Detailed information on overlap and assembly quality together with annotation of this entry is available at
 http://xgp.dna.affrc.go.jp/GenomeSeq.html.
 FEATURES
 Location/Qualifiers
 1..123673
 /organism="Oryza sativa (japonica cultivar-group)"

Percent Similarity: 97.84%	Conservative: 2
Best Local Similarity: 96.40%	Mismatches: 2
Query Match: 92.65%	Indels: 1
DB: 8	Gaps: 0
US-09-868-025-2 (1-143) x AP002837 (1-123673)	
QY 1 MetValAspThrAsn-PheProIleSerGluGlnSerGluThrHisAlaTrpCysTrpSer 20	
DB 57075 CCGGTAGATACAAATTTTCCATCATGAGCAATCAGAAACACATGCTTGGTGTGGAGC 57134	
QY 21 SerSerThrArgSerProSerArgHisHisLeuHisArgGluArgIleProCysLeu 40	
DB 57135 AGCAGCACCACAGAACGCGTCCGCCATCATCTTCATCGGCGGATCCCATGCCCT 57194	
QY 41 AlaLeuGlyValThrAlaIleCysSerLeuValTrpIleHisValSerHisGlyGly 60	
DB 57195 GCGTCGGTGCACGCAATTCAGTCTTCTTGGATACATGTCACCCAGCGTGTGA 57254	
QY 61 IleSerValAlaLeuCysSerGlnCysLeuGlnThrAsnAlaLeuArgProArgProAsp 80	
DB 57255 ATCAGCGTGCCTTTTGCAGCGAGTGCCTTCAGACGAATGCGTCCGCCCTCGACCCGAT 57314	
QY 81 CysLeuThrAsnAspGlyGlyCysTyrglyGluCysHisGlySerLeuGlyHisValAsp 100	
DB 57315 TGCTGACCAACAATGGAGGCTGCTATGTGTGAATGCCATGCGACTCTTGGGCGATGTTGAC 57374	
QY 101 ArgPheProGlnHisSerAsnGluTrpAsnSerGlyMetAspSerCysLysProLeuArg 120	
DB 57375 AGATTTCTTCAGCATTCATGATGAACAGCGCATGACAGCTCAACACCATCGAGG 57434	
QY 121 GlyGluPheLeuGlyValLeuThrProHisProLysMetGlu-PheAlaAlaIle 138	
DB 57435 CGAGATTTCTTGGTGTGCTCAGCGCACACCCAGAGGATGTTTCTGCCATT 57489	
RESULT 3	
AP003955/c	
LOCUS	80327 bp DNA linear HTG 21-MAR-2002
DEFINITION	Oryza sativa (japonica cultivar-group) chromosome 7 clone
ACCESSION	AP003955.1 GI:15021925
VERSION	HTG; HTGS_PHASE2.
KEYWORDS	Oryza sativa (japonica cultivar-group)
SOURCE	Oryza sativa (japonica cultivar-group)
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE	1
AUTHORS	Sasaki, T., Matsumoto, T. and Yamamoto, K.
TITLE	Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, BAC
JOURNAL	Published Only in Database (2001)
REFERENCE	2 (bases 1 to 80327)
AUTHORS	Sasaki, T., Matsumoto, T. and Yamamoto, K.
TITLE	Submitted (25-JUL-2001) Takuji Sasaki, National Institute of Agrobiological Resources, Rice Genome Research Program, Kannondai
JOURNAL	2-1-2, Tsukuba, Ibaraki 305-8602, Japan
COMMENT	(E-mail: tsasaki@ias.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/, Tel: 81-298-38-7441, Fax: 81-298-38-7468) The nucleotide sequence of this BAC clone was generated by combining Monoclonal and RFLP-Japan sequencing data. NOTE: It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have provided by the submitter. This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved. * NOTE: This is a 'working draft' sequence. * This sequence will be replaced * by the finished sequence as soon as it is available and * the accession number will be preserved.

FEATURES	Location/Qualifiers
source	1. 80327
	/organism="Oryza sativa (japonica cultivar-group)"
	/mol_type="genomic DNA"
	/cultivar="Nipponbare"
	/db_xref="taxon:39947"
	/chromosome="7"
	/clone="OJ1425_E04"
BASE COUNT	23838 a 16833 c 16495 g 23361 t
ORIGIN	
Alignment Scores:	
Pred. No.:	3,33e-45 Length: 80327
Score:	631.50 Matches: 117
Percent Similarity:	88.49% Conservative: 6
Best Local Similarity:	84.17% Mismatches: 13
Query Match:	78.64% Indels: 3
DB:	2 Gaps: 1
US-09-868-025-2 (1-143) x AP003955 (1-80327)	
QY 2 ValAspThrAsn-PheProIleSerGluGlnSerGluThrHisAlaTrpCysTrpSer 21	
DB 33724 ATAGACACAAATTTTCAAAGCAATGAACAATCAGAAACACATGCTTGGTGTGGAGC 33665	
QY 21 rSerThrArgSerProSerArgHisHisLeuHisArgGluArgIleProCysLeuAl 41	
DB 33664 CAAGCCACCAAGAGCCCGTCC---AGCCATCTTCATCGGAGCGGATCCCATGCTTGC 33608	
QY 41 aleuGlyValThrAlaIleCysSerLeuValTrpIleHisValSerHisGlyGly 61	
DB 33607 CCTCAGTGTCATGCAATTCAGGACTTGTGTTGGATCATGTCAGCCAGCGTGTGGAA 33548	
QY 61 eSerValAlaLeuCysSerGlnCysLeuGlnThrAsnAlaLeuArgProArgProAsp 81	
DB 33547 CGGCGTCGCGCTTTGCAGCCAAATGCTTCAGACGAATGCGCTCGGCCCTCGACCCGATTG 33488	
QY 81 eLeuThrAsnAspGlyGlyCysTyrglyGluCysHisGlySerLeuGlyHisValAsp 101	
DB 33487 TCTGGCCATATGAGAGGCTGCTATGTGTGAATGCCATGCGCATGTTGGCATGTTCCAG 33428	
QY 101 gPheProGlnHisSerAsnGluTrpAsnSerGlyMetAspSerCysLysProLeuArg 121	
DB 33427 ATTTCCTCAGCATTCGATGAATGGAACAGCGCATGACAGCTGCAACACCATCCCGGG 33368	
QY 121 yGluPheLeuGlyValLeuThrProHisProLysMetGlu-PheAlaAlaIle 138	
DB 33367 AGAATTTCTTGGTGTGCTCAGCCACACCCGAGAGGATGTTTGTGCCATT 33315	
RESULT 4	
AP003847	
LOCUS	191580 bp DNA linear PLN 10-APR-2003
DEFINITION	Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 7, BAC clone: OJ1714_H10.
ACCESSION	AP003847
VERSION	AP003847.3 GI:22831069
KEYWORDS	Oryza sativa (japonica cultivar-group)
SOURCE	Oryza sativa (japonica cultivar-group)
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE	1
AUTHORS	Sasaki, T., Matsumoto, T. and Yamamoto, K.
TITLE	Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, BAC
JOURNAL	Clone: OJ1714_H10
REFERENCE	2 (bases 1 to 191580)
AUTHORS	Sasaki, T., Matsumoto, T. and Yamamoto, K.
TITLE	Direct Submission
JOURNAL	Submitted (03-JUL-2001) Takuji Sasaki, National Institute of Agrobiological Sciences, Rice Genome Research Program, Kannondai
	2-1-2, Tsukuba, Ibaraki 305-8602, Japan


```

join(11092) .11178.12933. .13856)
  /gene="OJ1714 H10.5"
  /note="contains ESTs C19982(E11335), AU075687(E11335)
unknown protein"
/codon_start=1
/protein_id="BAC15932.1"
/db_xref="GI:22831070"
/translation="MAKKKPIVPMDISIGDEPEATPNMTWREREREGERSGSGGG
RVVRSKSLRFVTDGGRSPPVHTGFQDKDHAANAARACDGFACIIVCLLSVAD
BERSSELIIVPVPAGMDVDVTHLSPVSCDGTGRTTSVAGEEAVNVAHQAVTGG
VRRRLALRGSGGSHRRHAGARVPRRRERLGGVLAAGAGVLPVPRVPRRLQGA
PRAVQGGQTHDLKRLSPFSFAMLVLPDORRSGISGAQDHLLAGVVRPFPAGV
PRGCGSDDDLGSLPYQGYVPPPLKLVDFVADHLPAFFIVERRLSQISVAGHVLPDSE
EV"
join(114565) .14684.15475. .15609.16934. .16990)
  /gene="OJ1714 H10.6"
  /join(114565) .14684.15475. .15609.16934. .16990)
  /gene="OJ1714 H10.6"
  /note="This category is not included in IRGSP standard.
hypothetical ORF
predicted by GENSCAN"
19647. .20096
  /gene="OJ1714 H10.7"
19647. .20096
  /gene="OJ1714 H10.7"
  /note="hypothetical protein
predicted by GlimmerM etc."
/codon_start=1
/protein_id="BAC15933.1"
/db_xref="GI:22831071"
/translation="MQLTRGRFFCKIQLNTRRRGRRLVWHEETRPATGTRCLGW
RCQVTRLAWSDQAARSLASREDAWPAETAGTMRGSGAGCGGAGYARRLAC
GHRCSGSPCCQRLSGRSGRQHGFCFAGGERVKIQGLSRAGNDDA"
join(22191) .22749.22910. .22914)
  /gene="OJ1714 H10.8"

```

CDS

```

/genes="Q9J714_H10.8"
/notes="hypothetical protein
predicted by GlimmerM etc."
/codon_start=1
/protein_id="BAC15934.1"
/acc_xref="GI:29831072."
/translations="MTKPGCLARRSAYVTSRQIPTRSQLMCAGGQWPPVVDGGG
SYVYKSGNAAVIVCLCAVAAYFLMAGITITQCLCKRPPAPAPQPPRRDRDDDD
IVAVGDSGTAQLPVLVYVGSARTQIAGSSVGGGVDVNLALPECGHLTHROCI

```

```

complement(join(24053).
/genes="OJ1714_H10.9".
/notes="This category is not included in ICGSP standard.
complement(join(24053).
/genes="OJ1714_H10.9".
/notes="This category is not included in ICGSP standard.

```

hypothetical ORF	
predicted by GENSCAN ^a	
complement [join (25444,	.26085,26479. .26805,26890. .27060)
/gene="O01714.H10.10"	
complement [join (25444,	.26085,26479. .26805,26890. .27060)

```
similar to Oryza sativa chromosome10, OSUBA0056G17.7
unknown protein"
/codon start=1
/protein id="BAC15935.1"
/db xref="GI:22831073".
/translators="M17MGFSGRMEKFSSFLNTEASSENUGPNTVERCPPLNNG
```

TDANPQILFNP,LAARAAT,ISLSA,FGFPGFPNFGKGRKQKNNLNNGKPNKPN,
 NNSMPOKGGNSSHSHEANWENMGQCP,LARGYRAMSGJLPI,VAKALQPPAGVKLK,
 PPAVVAI,PAALARTKLSVRPOTPLPAKXVAI,ALLGMAANI,ELGVREHTKFKSPQMF,
 AAHAANVPF,IGHKRLKSVNPKP,LAFTNAFTI,NAASIIQGTISRAEIRUKALAKGDADST,
 TVADNTPKNSGNSDTEGKAWDP,PLANKMWRASGSGVAAPTPSNCF".

gene

```
/gene="OJ1714_H10.11"  
30681. 31229
```

```

FEATURES
  source
    1..1164
      Location/Qualifiers
        /organism="Oryza sativa"
        /mol_type="genomic DNA"
        /db_xref="taxon:4530"
  BASE COUNT      294 a      286 c      297 g      287 t
  ORIGIN

Alignment Scores:
  pred. No.:      3e-45
  Score:          610.50
  Percent Similarity: 89.84%
  Best Local Similarity: 86.72%
  Query Match:    76.03%
                  6
  DB:
  Gaps:          1
  Indels:        2
  Mismatches:    11
  Conservative:  4
  Matches:       111
  Length:       1164

```

```

08-09-888-025-2 (1-143) X AK653397 (1-1164)

QY 12 SerGluThrHisAlaTrpCysIrpSerSerThrThrArgSerProSerArgHis 31
DB 1164 TCAGAAACACATGTTGGTGTGGTCAGCAACGCCACCAAGACCGCGTC ---AGGCAT 1108
QY 32 LeuHisArgGluArgIleProCysLeuAlaLeuGlyValThrAlaIleCysSerLeuVal 51
DB 1107 CTTTCATCGGAGCGGATCCCATCCCTTGGCCCTCAGTGTCACTGCAATGCCAGACTTGT 1048
QY 52 TrpIleHisValSerHisGlyGlyIlyIleSerValAlaLeuCysSerGlnCysLeuGln 71
DB 1047 TGGATATCATGTACAGCCACGGTGCTGGAAATCGGCGTGGCCCTTTGCAGCAATGCCCTTCAG 988
QY 72 ThrAsnAlaLeuArgProArgProAspCysLeuThrAsnAsnGlyGlyCysArgGlyGlu 91
DB 987 ACGAATGCGCTCGGCCCTCGACCCGATTGCTGCGCAATAATGAGAGGTGCTATGGTGAA 928
QY 92 CysHisGlySerLeuGlyHisValAspArgPheProGlnHisSerAsnGluTrpAsnSer 111
DB 927 TGCCATGGCAGTCTTGGCATGTTACAGATTTCCTCAGCATTCGATGAATGGAACAGC 868
QY 112 GlyMetAspSerCysIlyProLeuArgGlyGluPheLeuGlyValLeuThrProHisPro 131
DB 867 GGCATGACAGCTGCAACCACTGCGGGGAGAAATTCCTTGGTGTGCTCACGCCACACCCC 808
QY 132 LysMetGlu-PheAlaAlaIle 138
DB 807 GAGAGGGATGTTGCTGCCATT 786

RESULT 5
LOCUS AK554660/c
DEFINITION Sequence 4530 from Patent WO03000898.
ACCESSION AK554660
VERSION AK554660.1 GI:29157474
KEYWORDS
SOURCE
ORGANISM
Oryza sativa
Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

```

REFERENCE
AUTHORS

Ehrhartoideae; Oryzeae; Oryza.

REFERENCE
TITLES

Chang, H.S., Chen, W., Cooper, B., Glazebrook, J., Goff, S.A., Hou, Y.M., Katagiri, P., Quan, S., Tao, Y., Whitham, S., Xie, Z., Zhu, T. and Zou, G.
Plant genes involved in defense against pathogens
Patent: WO 0300898-A 4530 03-JAN-2003;
Syngenta Participations AG (CH)

JOURNAL

FEATURES

Location/Qualifiers
1..1104
/organism="Oryza sativa"
/mol_type="genomic DNA"
/db_xref="taxon:4530"
278 a 268 c 281 g 274 t 3 others

BASE COUNT 278 a 268 c 281 g 274 t 3 others

ORIGIN

Alignment Scores:

Pred. No.: 0.000364 Length: 1104
Score: 148.50 Matches: 46
Percent Similarity: 49.58% Conservative: 13
Best Local Similarity: 38.66% Mismatches: 40
Query Match: 18.49% Indels: 20
DB: 6 Gaps: 4

US-09-868-025-2 (1-143) x AX654660 (1-1104)

Qy 35 GluArgileProCysLeuAlaLeuGlyValThrAlaileCysSerLeuValTrpIleHis 54
Db 1041 AAGCGGATCCAGAACTGGACATC-----ATCAGT 1012
Qy 55 ValSerHisGly-----GlyGlyLeSerValAlaLeuCysSerGln----- 68
Db 1011 GTATCTCCCGTTTTCAGGCTCATTCTGCTAGCATCACCCGTGCTACAGCAGCAAC 952
Qy 69 -----Cys-----LeuGlnThrAsnAlaLeuArgProArgProArgCysLeuThrAsnAsn 85
Db 951 CTTTGTGCGCTCTTCATCTGATGATGATGCTAGCTCTGACCAATTCGTGACCAATAT 892
Qy 86 GlyGlyCysThrGlyGlyCysHisGlySerLeuGlyHisValAspArgPheProGlnHis 105
Db 891 TGAGGCGAGCTATGTGTAAGCGGCATGGCTGTCTTTGGCATCAGCAGACATCTCCAGCAT 832
Qy 106 SerAsnGluThrAsnSerGlyMetAspSerCysLysProLeuArgGlyGluPheLeuGly 125
Db 831 GCCTATGAAGTACCGCGCATGCTGCTGGCGAACCACCTGCACTGAGATTTCTCGT 772
Qy 126 ValLeuThrProHis-ProLysMetGluPheAlaAlaileArgAlaGlyLysVal 143
Db 771 GTGCTCCCTCCAGATGCCAAGCGGAACGTTTGTGCTGATACCAAGCAACGCCATC 717

RESULT 7

AE017123

LOCUS

AE017123 95893 bp DNA linear PLN 06-JUN-2003
Oryza sativa (japonica cultivar-group) chromosome 10, section 77 of 77 of the complete sequence.

ACCESSION

VERSION

AE017123.1

KEYWORDS

SOURCE

ORGANISM

Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
1 (bases 1 to 95893)

REFERENCE

AUTHORS

CONSRM

TITLE

The Rice Chromosome 10 Sequencing Consortium
In-depth view of structure, activity, and evolution of rice
chromosome 10
Science 300, 1566-1569 (2003)
2 (bases 1 to 95893)
Buell, C.R., Wing, R.A., McCombie, W.R., Messing, J. and Yuan, Q.
Direct Submission
Submitted (05-MAY-2003) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA

COMMENT

This is the pseudomolecule for rice chromosome 10, which was constructed by resolving discrepancies between overlapping BACs, trimming the overlap regions, and linking the unique sequences to form a contiguous sequence. Genes in individual BAC clone were identified by a combination of several methods: Gene prediction programs, searches of the complete sequence against a peptide database and EST databases. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>).

FEATURES

source

1..95893
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="10"

misc_feature

<1..95317
/note="Chromosome Sequence Derivation: nucleotide sequence in this region was derived from BAC clone OSUNBA0056G17 (GB:AC018727)."
complement(<1834..>3425)
/locus_tag="OSUNBA0056G17.12"
/note="EST D15631, C28081, C93504 from this gene"
complement(join(<1834..1923,2955..3039,3122..>3425))
/locus_tag="OSUNBA0056G17.12"
complement(join(1834..1923,2955..3039,3122..3417))
/locus_tag="OSUNBA0056G17.12"
/codon_start=1
/product="unknown protein"
/protein_id="AAP55179.1"
/db_xref="GI:31433700"

gene

/translation="WRRWCNCAAGLGRLLSSSVSAPARPLPAHLIPSPRLPSS
RHLLTTPGLHPSPPPMQMOTQVHRFAADRGRAPRTPTTSVKYKIKPPSSMK
PFRVMDQGVRRWRAGKEHNAHLSKEAKRLRKPALVLAAYKVIKXNFC"

mRNA

<4368..>7204

CDS

/locus_tag="OSUNBA0056G17.8"
/note="tag similar to peptide transporter AtPTR2 B GB:AA00858
GI:633940 (Arabidopsis thaliana); EST AU068484 from this
gene"
join(4368..4453,4697..4763,5040..5257,5412..5968,
6385..>7204)
/locus_tag="OSUNBA0056G17.8"
join(4406..4453,4697..4763,5040..5257,5412..5968,
6385..7204)

gene

/codon_start=1
/product="putative peptide transporter"
/protein_id="AAP55180.1"
/db_xref="GI:31433701"

mRNA

/translation="MGMSGDAQPLLLHQASASNQHYTKPPPNWKAPALILAFEPLE
SIAYAGISLNLVVLGTVLHGTAGNANVTWNGTTLTPVLGFLADTYMGKYTI
AISTVPLGLLIVTASAVIPSLPAPCGNGSCPRATGQYVFTVTLVLSGTGGV
KSLALLPFGADYNDNLSLEKKSQSFSLPFTAINLGVISCTVVMIOONVMSLGF
GSSCLLVATVAFVLAFTVYKQLPSPGSLKSVYVASFVKRKLVPADNALLYE
GDADLSNQSVLAHTDGRWLDKAAVFEVEBEINKDEGGGGLQCSVTQVEEVK
ILLRLPIWTSVLYAASLGQTATTFVQGNANTKIGSFVPAASLNSARVIFMMIV
VFQDTVVVPIARVTGNPAGLTQLRMGVGRLLAVPALVAALVETWRLSRVDDGN
LSIAWLPQFVILACSDVFCGIAOLEFFYSAPVSMRSLCSAFSLALSGLGYVNSLV
VSIIVAVTTTSNKGWLPADLNDGLDYYTFLWTGISAINFVYVYAPAKNYTVKRLAVP
HSH"

CDS

complement(<7577..>12536)
/locus_tag="OSUNBA0056G17.19"
/note="tag similar to polyprotein GB:AAD13304 GI:4235644
(lycopersicon esculentum)"
complement(join(<7577..7689,7783..7959,8059..8184,
10478..10687,10865..10995,11095..11252,11633..11691,
12215..>12536))
/locus_tag="OSUNBA0056G17.19"

gene

complement(<7577..>12536)
/locus_tag="OSUNBA0056G17.19"
/note="tag similar to polyprotein GB:AAD13304 GI:4235644
(lycopersicon esculentum)"
complement(join(<7577..7689,7783..7959,8059..8184,
10478..10687,10865..10995,11095..11252,11633..11691,
12215..>12536))
/locus_tag="OSUNBA0056G17.19"

mRNA

complement(join(<7577..7689,7783..7959,8059..8184,
10478..10687,10865..10995,11095..11252,11633..11691,
12215..>12536))
/locus_tag="OSUNBA0056G17.19"

[illegible]


```
/db_xref="GI:12039385"
/translation="MAVDDLNHHLLIANGVGEBAATACGCLRTLVIRNMAQVLSGLIV
AAKLVQATPAASAKLFPFALLFPFSLANGFNAPFDPATVFIHRWLP
EPKVPSPGLPFWAAVFAVAVYPTALGTATTCCLPFLASLTVFQYLTPA
LPSGVKVLHPICNCKSLAAVAYGLSRGVDALVDYLTPSPNPGAGDILMG
PLSGVITSPFSPNCKSLAAVAYGLSRGVDALVDYLTPSPNPGAGDILMG
ILPRCTIVALLAVVSLFEGVNTSVTAUVVLGLGANFAQAVMDKRLKQPIARGI
GTASRAGLGTAAVSAKEPALPCALAYALTGVVASLFCSLPAIRHSLVFIAGDASA
SQTHFSY"
complement(10278..10350)
/rpt_family="AT rich"
complement(12207..12227)
/rpt_family="AT rich"
complement(14187..16368)
/gene="OSUNBA0056G17.15"
/notes="highly similar to cholinephosphate
cytidylyltransferase GB:CAA70317 GI:1657382 (Pisum
sativum); EST C26593 from this gene"
complement(join(14187..14282,14405..14515,14635..14703,
14841..14996,15085..15127,15235..15330,15805..15918,
16059..16368))
/gene="OSUNBA0056G17.15"
complement(join(14187..14282,14405..14515,14635..14703,
14841..14996,15085..15127,15235..15330,15805..15918,
16059..16243))
/gene="OSUNBA0056G17.15"
/codon_start=1
/product="putative cholinephosphate cytidylyltransferase"
/protein_id="AAG46173.1"
/db_xref="GI:12039387"
/translation="MARVSHSRKSTRAPSNSKPTSTOTPPPLTDRPVVADGI
FDLFFHGARALEQALFFNTYLLVCCNDELNRYKGVTVQDRYSLSLHCKVW
DEVIPDAWLTQSDIDKHQDYVAHDALPDATSGAANDVIFVKIGKIFKTKRID
GVSTDLMTLKQDYNLARGYTRKOLGVSYKKEKOLQVNMINKLRETVRAH
CEKLTQVAKTAGIINHEELANADRWAGVLEKFEHCHNEMTAIKSRIOEKLRQTSR
GIIGLMQOPVAA"
complement(16595..16625)
/rpt_family="TAA n"
complement(16673..16698)
/rpt_family="(CATCG)n"
complement(17509..20874)
/gene="OSUNBA0056G17.30"
/notes="similar to 4 coumarate CoA ligase GB:AAC24503
GI:3258635 (Populus tremuloides)"
complement(join(17509..17640,18143..18245,18370..18437,
18565..18722,19375..19564,19831..20874))
/gene="OSUNBA0056G17.30"
complement(join(17509..17640,18143..18245,18370..18437,
18565..18722,19375..19564,19831..20874))

Alignment Scores:
Pred. No.: 0.0563 Length: 139999
Score: 148.50 Matches: 46
Percent Similarity: 49.58% Conservative: 13
Best Local Similarity: 38.66% Mismatches: 40
Query Match: 18.49% Indels: 20
DB: 8 Gaps: 4

US-09-868-025-2 (1-143) x AC018727 (1-139999)

QY 35 GluArgileProcysleuAlaLeuGlyValThrAlaIleCysSerLeuValTrpIleHis 54
DB 126973 AAGCGGATCCCAAGACTGCATC-----ATCAGT 127002
QY 55 ValSerHisGly-----GlyGlyIleSerValAlaLeuCysSerGln----- 68
DB 127003 GTATCTCCCGGTTTTCAGGCTCATTCGTGCTAGCATCACCTGCCTACCAGCAGCAAC 127062
QY 69 -----Cys-----LeuGlnThrAsnAlaLeuArgProAspCysLeuThrAsnAsn 85
DB 127063 CTTTGCTGCCCTCTCAATCTGATACGCTCAGCTCTCGAACAACAAATGTCTGACCAATAT 127122
```

```
QY 86 GlyGlyCysValTrpGlyCysHisGlySerLeuGlyHisValAspArgPheProGlnHis 105
DB 127123 TGAGGAGCTATGGTAGCGCATGGCTGCTTTGGCATCAGCAGACTCTCTGAGCAT 127182
QY 106 SerAenGluTrpAsnSerGlyMetAspSerCysLysProLeuArgGlyGlyLysVal 125
DB 127183 GCTTATGAAGGTACCGCGCATGAGCTGGCGCAACCACTGCATCAGAGAAATTTCTCCCT 127242
QY 126 ValLeuThrProHis-ProLysMetGluPheAlaAlaIleArgAlaGlyLysVal 143
DB 127243 GTGCTCTCCAGTCCAGTCCAGGGAGCGTTCCTGCTCCATACCAGCAAGCCCATC 127297

RESULT 9
BT002755/c 1060 bp mRNA linear PLN 15-JAN-2003
LOCUS Arabidopsis thaliana clone C104940 unknown protein (At5g45410)
DEFINITION mRNA, complete cds.
ACCESSION BT002755
VERSION BT002755.1 GI:27754260
SOURCE FLI CDNA.
ORGANISM Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; euroside II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE 1 (bases 1 to 1060)
AUTHORS Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,
Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C.,
Yu,G., Yuan,S., Chen,H., Cheuk,R., Jones,T., Kim,C.J., Nguyen,M.,
Palm,C.J., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Davis,R.W.,
Ecker,J.R. and Theologis,A.
TITLE Arabidopsis Open Reading Frame (ORF) Clones
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1060)
AUTHORS Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,
Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C.,
Yu,G., Yuan,S., Chen,H., Cheuk,R., Jones,T., Kim,C.J., Nguyen,M.,
Palm,C.J., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Davis,R.W.,
Ecker,J.R. and Theologis,A.
TITLE Direct Submission
JOURNAL Submitted (15-JAN-2003) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
COMMENT Annotation based on July 2002 version of the Arabidopsis genome
submitted to Genbank.
FEATURES
Location/Qualifiers
source 1..1060
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="5"
/clone="C104940"
/ecotype="This clone is in pUNI 51."
/ecotype="Columbia"
1..1060
/gene="At5g45410"
1..1029
/gene="At5g45410"
/codon_start=1
/evidence=experimental
/product="unknown protein"
/protein_id="AA02584.1"
/db_xref="GI:27754261"
/translation="MDQEGPFAAEIIOKPPFLRNINKPTNLSFSLSPFPVQGGKG
PIEDPSPDASAKLPHKGIIVLSPGADDEBAGRALQFNPLAGKVATISLGA
FGPGPGFPGFSEKWKQKQKPKSPKQSGSDSKHEAVGDEWLKTCNCFKSPAA
SKYPTISKALTLPQMKYRCAPIVAAALSKALVLSRPOPJPERMLALQGM
AANVLSGWRTEKTFSPAMFVAAVPIAMLRKSLVLMPTKALTIGASILGOVI
GSAERYRLKAVAKWVPVMTVMVSYNQSPGSDGISGCHGIGEGVKQVYNNYNGIS
AKSTGLCY"
misc_difference 375
/gene="At5g45410"
/notes="compared to genomic sequence"
/replace="a"
```



```

QY 127 LeuthProHIS-ProLysMetGlupheAalAaIaIle 138
Db :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
656 TTTCTCCATACACCTAGACGCACATTTCGACGCAAT 620

RESULT 11
LOCUS AB018113 85020 bp DNA linear PLN 27-DEC-2000
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 5, F1 clone:MFC19.
ACCESSION AB018113 BA000015
VERSION AB018113.1 GI:3702731
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (sites)
Kaneko,T., Katoh,T., Sato,S., Nakamura,Y., Aemizu,E., Kotani,H.,
Miyajima,N. and Tabata,S.
Structural analysis of Arabidopsis thaliana chromosome 5. IX.
Sequence features of the regions of 1,011,550 bp covered by
seventeen P1 and TAC clones
DNA Res. 6 (3), 183-195 (1999)
2 (bases 1 to 85020)
Nakamura,Y.
Direct Submission
Submitted (06-OCT-1998) Yasukazu Nakamura, Kazusa DNA Research
Institute, Department of Plant Gene Research, 1532-3, Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail:yakam@kazusa.or.jp,
Tel.81-438-52-3935, Fax:81-438-52-3934)
Address for correspondence: kaos@kazusa.or.jp
For the latest information on annotation of this clone, please see
http://www.kazusa.or.jp/kaos/cgi-bin/agg_graph.cgi?cm=MFC19
Genes with similarity to proteins in the databases are described in
'product' or 'note' qualifiers. Genes that have no significant
protein similarity are described as 'unknown protein'.
The software programs used to predict genes include: Orail
(Informatics Group, Oak Ridge National Laboratory,
http://compbio.ornl.gov/Grail-1.3/),
GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
SplicePredictor (Volker Brendel, Stanford University,
http://gremlini.zoool.iastate.edu/cgi-bin/sp.cgi).
Genes encoding tRNAs are predicted by tRNAscan-SE
(Sean Eddy, Washington University School of Medicine, St. Louis,
http://genome.wustl.edu/eddy/tRNAscan-SE/).
This sequence may not be the entire insert of this clone. It may be
shorter because we remove overlaps between neighboring submissions.
The 5' clone is K9E15 and the 3' clone is K2N11.
Location/Qualifiers
1. 85020
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
/clone="MFC19"
/clone_lib="Mitsui P1"
complement(join(1506..1744,1966..2260))
/note="gene id:MFC19.1
unknown protein"
/codon_start=1
/evidence=not experimental
/product="replication protein A1-like"
/db_xref="GI:9758725"
/translation="MGNDNDKDGPHGYPPGAYPPGAYPPGAYPPGAYPPGAYPPGAY
PPGAYPPGAYPPGAYPPGAYPPGAYPPGAYPPGAYPPGAYPPGAYPPGAYPPGAYPPGAY
GGGAGAAAGYAGYVHSHSGPYGHAAYGHGFGHGYGCHGHGKFKGKFKGK
KGGHGMFGGKFKKWK"
complement(join(4189..4259,4368..4544,4628..4802,
CDS
4887..4997,5148..5223,5337..5410,5493..5696,6014..6076))
/note="gene id:MFC19.2
unknown protein"
/codon_start=1
/evidence=not experimental
/product="ATP-dependent Clp protease-like protein"
/db_xref="GI:9758726"
/translation="MSVSDDEDCFAFLPSESVSSVEDKRETEKPEAKRQRIEKGTRK
ALEKDEQKNGKNDKADKIRKQVNTVPTLFRHLKFLSSDLVSCSLCKEFLN
FAADESLMRRLYCIKRGTLPRKLRSAWKLYLDRDEQMIKLVTRTCPSDFREY
VHMAAKRSQAPLPSQWDDRIILDNITLVKQVSLMKSGKLTOKATYGHICGTKCSY
HQIDDPFICKETGNVHVCDNCKEVLSPGDLNVCCTISQVCSDDLVTETPDAGCY
EEAELEAVFTDKSLRLARAPELGYNCCDQELELTLRF"
join(7817..8031,8442..8558,8650..9091,9265..9416,
9532..9688)
/note="gene id:MFC19.4"
/codon_start=1
/evidence=not experimental
/product="modulin-like protein"
/db_xref="GI:9758727"
/translation="MAAPAILGGDATERETMAHSMTLVQVINGYHVYTKVALNV
GVNLVPCVFRDLALLSILAPAFPRFRIFGNOLLFLMGLSYTNPYAAAIQPSIVF
TELLAVLMTKVLNLLKVEGTGKGLVGVGAIAIAPALPGALPGADAAOVKSV
IIDRSQPELNGMLVSSFLGFGDQWHLGLCLIGNCMMAAFILAVOAPVLKYPAYLS
VARYSFFGASIMITAILFYREPKDMSLTQSEVLAVIAGVFASALNGLTWSNKL
LGALVSLYNPQIPATSAFLSTIFLSPFYLGSVLGAILICGLYMTWASYREQQT
SAGNIASSDVRISSEPIFYRDETK"
join(10743..10910,11288..11395,11531..11626,11725..11807,
12163..12381,12612..12874,13160..13398,13823..14253,
14382..14585,14831..15102)
/note="gene id:MFC19.5"
/codon_start=1
/evidence=not experimental
/product="urea active transporter-like protein"
/db_xref="GI:9758728"
/translation="MATCPFFSTKYDGGCGQROSSFGGTTLDQGVYAVILG
SNAPFAVPTSELVLEKRYGARTSFMFNAGRNKVIAGLSYTNPTVAAAIQPSIVF
GVAVGVSGPFPYATGATOVLLFGMAIEIKKAPNAFTVCEIVKRWGATATHIV
FLVPLCLANVVVATMLLGGSAVNALTVNLVYASFLIPLGVVVVTLAGLKAFLA
SVHSYVHVVALVVFVLYVTSKELASPVVYDLKDMVKMSTETESHGQACG
PVDMGFRSGY.TMSSGGAVFLGIVNFGTVDMGVWVSAIARPPSTHKGYLIG
GLVMPVPSLATSLGICALALDLPISKDEADRLVPPATAIALMGSGSLTLTWLF
MAVTSAGSELAVSSLTVDYIYTYINPRATGRQILKISCAVLGFCFPMGILAVL
NKVAGSLGNVLMQVTLGSAVILPAMLLMSKANAGAILGATSGCVGFIITLDTA
KCYGRVLDSTGKGMPLAGNVAILTRPNYDNSTREIKUVEYVASSGSDQVDP
AELEKELRRKAKAWIKVGLVFTLLIYVIMPVLSLPARVPSRGRYFWFVAIVAGT
IGSVILGLPLVESWDTIKVCMGMFTNDRVKKLDDLRRLRALTMVAPAEKIVYL
ELETKQNDDEG"
join(15757..16151,16482..16992)
/note="gene id:MFC19.6"
/codon_start=1
/evidence=not experimental
/product="ATP-dependent Clp protease-like protein"
/db_xref="GI:9758729"
/translation="MGTLSLSSKPSLVSRNLSSSSASSSSFPKNNYLKPTKLI
SPLTRPSPPLRFANAKISQTESAIGAESDVNGLLRIRVFLGSSIDOFVAD
AIMSQLLDKADPKKIKLFINFGSGSLNLYVWFSATMAIYVQVLYRADVSTI
ALGIAASTAIIAGTAKGKFAFNPTRIMIHOPLGASQAIIVETQAEVMMHKN
VTSIAGCTSRSEFQVLDIDRYMSPIEAVEYGLIDGVIDGSIILPLEPFDVXP
RVNVEISKDPMKPLTPEIDDEIY"
join(18396..18688,18782..21050)
/note="gene id:MFC19.7"
/codon_start=1
/evidence=not experimental
/product="replication protein A1-like"
/db_xref="GI:9758730"
/translation="MAVSLTEGVMMKMLNGEVTSETDMPVLTQELKQSLKHQK
ESSNYKFLSDGTDLAAGMLTSLNSLVNQTIGLSVRLTHYICNLITQIRIVVI
MQLEVIVEKCNIIIGNPKSGHNPQRGVNTQSGSGSQQQARRSDVNGRYGVA

```


----- Genome Center -----					
Center: Washington University Genome Sequencing Center					
Center code: WUGSC					
Web site: http://genome.wustl.edu/gsc/index.shtml					
Contact: submissions@wustl.wustl.edu					
----- Project Information -----					
Center project name: M.BA0480508					
----- Location/Qualifiers -----					
1..206926					
/organism="Mus musculus"					
/mol_type="genomic DNA"					
/db_xref="taxon:10090"					
/chromosome="5"					
/clone="RP23-480E8"					
BASE COUNT 67331 a 37362 c 37919 g 64314 t					
ORIGIN					
Alignment Scores:					
Pred. No.:	461	Length:	206926		
Score:	106.50	Matches:	36		
Percent Similarity:	40.00%	Conservative:	16		
Best Local Similarity:	27.69%	Mismatches:	61		
FEATURES					
source					
174874					
Length:	174874				
Pred. No.:	387	Matches:	36		
Score:	106.50	Conservative:	16		
Percent Similarity:	40.00%	Mismatches:	61		
Best Local Similarity:	27.69%	Indels:	17		
Query Match:	13.26%	Gaps:	4		
DB:	10				
US-09-868-025-2 (1-143) x AC123054 (1-174874)					
Qy	9	SerGluInSerGlutThrHisAlatPCysTrpSerSerThrTrArGser-----	26		
Db	27833	TCCAGCAGGACTGAATGTCTCCCTCTGCTTCTTCGTACACCACAGATACCAAGAA	27892		
Qy	27	-----ProSerArgHisHisLeuHisArgGluArgIleProCysLeuAlaLeuGly	43		
Db	27893	CATCTGAGGCATGTTCCAGCACACCTCGCGCCTGTCTCAGTGCTGTGTGTCATGTCAGC	27952		
Qy	44	valThrAlarIeCysSerLeuValTrpIleHisValSerHisGlyGlyIleSerVal	63		

```
Query Match: 13.264 Indels: 17
DB: 10 Gaps: 4
US-09-868-025-2 (1-143) x AC124497 (1-206926)
QY 9 SerGluGlnSerGluThrHisAlaTrpCysTrpSerSerThrArgSer----- 26
Db 5516 TCAGCAGGACCTGAATGTCTCCCTCTGCTCTTCTTGACACACAGATACCAGAA 5457
QY 27 -----ProSerArgHisHisLeuHisArgGluArgIleProCysLeuAlaLeuGly 43
Db 5456 CATCTGGAGCCATGTTACACACACCTGGCGCTGTGCAGGTGGTCTGTGTTTCATGTTTCAGC 5397
QY 44 ValThrAlaIleCysSerLeuValTrpIleHisValSerHisGlyGlyGlyGlyVal 63
Db 5396 ACACCTGGCGGCTGTGCAGGTGGTCTGTGTTTCATGTTTCAGCACACCTGGCGGCTGTGAGGT 5337
QY 64 AlaLeuCysSerGlnCysLeuGlnThrAsnAlaLeuArgProArgProAspCysLeuThr 83
Db 5336 GGTCTGTGTTCATGCTACACACCTGGCGGCTGTGCAGGTGGTCTGTGTTTCATGCTCAGC 5277
QY 84 AsnAsnGlyGlyCysTyrglyGlyCysHisGlySerLeuGlyHisValAspArgPhePro 103
Db 5276 ACACCTGGCGGCTGTGCAGGTGGTCTGT-----GTTTCATGTT----- 5241
QY 104 GlnHisSerAsnGluTrpAsnSerGlyMetAspSerCys-----LysProLeu 119
Db 5240 CAGCACACCTGTGACCTGTGCAGGTGGTCTGTGTTTCATGTTTCAGCACACCTGGCGGCTGTC 5181
QY 120 ArgGlyGluPheLeuGlyValLeuThrPro 129
Db 5180 AGGTGGTCTGTGTTTCATGTTTCAGCACACCT 5151
```

Search completed: November 14, 2003, 13:11:39
Job time : 2264 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2003, 07:46:51 ; Search time 1939 Seconds
(without alignments)
8435.746 Million cell updates/sec

Title: US-09-868-025-1

Perfect score: 673
Sequence: 1 tttaactgtcctgtcgtgat.....actgtgttaaaaaaaaaa 673

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST.*

1: em_estba.*

2: em_esthum.*

3: em_estin.*

4: em_estmv.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_hic.*

9: gb_esti.*

10: gb_estc.*

11: gb_hic.*

12: gb_est3.*

13: gb_est4.*

14: gb_est5.*

15: em_estfun.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inh.*

19: em_gss_pln.*

20: em_gss_vrt.*

21: em_gss_fun.*

22: em_gss_mam.*

23: em_gss_mus.*

24: em_gss_pro.*

25: em_gss_rnd.*

26: em_gss_pmg.*

27: em_gss_vrl.*

28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	631.2	94.1	780	14	CB655151
C 2	632.8	94.0	777	14	CB627453
C 3	630.6	93.7	723	9	AU068433
C 4	630.6	93.7	776	14	CB000367

5	628.6	93.4	705	9	AU165747
6	628.6	93.4	712	9	AU075373
C 7	628.6	93.4	805	14	CB660999
C 8	627.4	93.2	582	14	CB647623
C 9	627.4	93.2	763	14	CB663774
C 10	619.4	92.0	768	14	CB649882
C 11	614.4	91.3	792	14	CB647260
C 12	613.8	91.2	727	14	CB655085
C 13	600.8	89.3	718	9	AU031690
C 14	573.6	85.2	622	13	BO906858
C 15	565.6	84.0	850	14	CB619892
C 16	560.6	83.3	916	14	CB633818
C 17	546	81.4	799	14	CB620522
C 18	540.8	80.4	836	14	CB659692
C 19	537.6	79.9	653	14	CB643278
C 20	467.2	69.4	486	9	AT003421
C 21	438.8	65.2	510	12	BM419698
C 22	424.6	63.1	504	12	BI811292
C 23	420.8	62.5	686	9	AU165429
C 24	403.2	59.9	459	9	AT003640
C 25	384.2	57.1	449	9	AT003707
C 26	361.4	53.7	378	9	AU093781
C 27	359	53.3	514	12	BI809607
C 28	354.8	52.7	419	14	D48832
C 29	349.4	51.9	691	10	BE229010
C 30	333.2	49.5	484	12	BI797263
C 31	320.8	47.7	382	9	AU030190
C 32	316.2	47.0	479	12	BI809674
C 33	310	46.1	591	14	CA759397
C 34	300.6	44.7	410	12	BM420265
C 35	288.8	42.9	535	10	BE229158
C 36	271.6	40.4	464	9	AU184049
C 37	265.8	39.5	321	10	BE230720
C 38	258.4	38.4	782	28	A2133549
C 39	247.6	36.8	615	13	BQ172639
C 40	246	36.6	703	12	BM078268
C 41	245	36.4	560	9	AI737464
C 42	244	36.3	670	14	CD052275
C 43	241.2	35.8	530	14	CA617150
C 44	230.4	34.2	275	12	BI809342
C 45	222	33.0	557	12	BJ322180

ALIGNMENTS

RESULT 1	CB655151/c	780 bp	mRNA	linear	EST 09-APR-2003
LOCUS	OSJNEC08E12.1	OSJNEC	Oryza sativa	(japonica cultivar-group)	cdna
DEFINITION	Clone OSJNEC08E12.1	3', mRNA sequence.			
ACCESSION	CB655151				
VERSION	CB655151.1	GI:29658876			
KEYWORDS	EST.				
SOURCE	Oryza sativa (japonica cultivar-group)				
ORGANISM	Oryza sativa (japonica cultivar-group)				
REFERENCE	1 (bases 1 to 780)				
AUTHORS	Jantasuriyarat, C., Lu, G., Gowda, M., Hatfield, J., Zhou, B., Mazur, E., Kudrna, D., Dean, R., Soderlund, C., Wing, R. and Wang, G.				
TITLE	Large-scale identification of ESTs involved in the interaction between rice and Magnaporthe grisea				
JOURNAL	Unpublished				
COMMENT	Contact: Rod Wing Arizona Genomics Institute University of Arizona Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ 85721-0088, USA Tel: 520 626 3967 Fax: 520 621 3288 Email: http://genome.arizona.edu				

PCR Primers

FORWARD: gta aaa cga cgg cca gtc
 BACKWARD: gga aac agc tat gac cat g
 Plate: 08 row: E column: 12
 Seq primer: gga aac agc tat gac cat g.

FEATURES

Location/Qualifiers
 1..780
 /organism="Oryza sativa (japonica cultivar-group)"

/mol_type="mRNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="OSUNEC08812"
 /tissue_type="leaf"
 /dev_stage="3 week"
 /lab_host="DH10B"
 /clone_lib="OSJNEC"
 /note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:
 XhoI; 6 hrs after inoculation with Rice Blast (C9240-1)"

BASE COUNT 213 a 206 c 176 g 185 t
 ORIGIN
 Query Match 94.1%; Score 633.2; DB 14; Length 780;
 Best Local Similarity 98.6%; Pred. No. 2.2e-155;
 Matches 649; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

QY 17 GGATGGCAGCAAC-TCCATCTTGGGTGTGGCTGAGCACACCAAGAAATTCCTCCCTC 75
 Db 695 GAATGGCAGCAACATCCCTCTTGGGTGTGGCTGAGCACACCAAGAAATTCCTCCCTC 636
 QY 76 AGTGGTTTGACGTGTCCATGCCGTGTTCATTATTTGGAATCTGAGGAATCTGTCA 135
 Db 635 AGTGGTTTGACGTGTCCATGCCGTGTTCATTATTTGGAATCTGAGGAATCTGTCA 576
 QY 136 ACATGCCAAGACTGCCATGCATCACCATAGAGCTCCATTTGTTGGTCAGACATCG 195
 Db 575 ACATGCCAAGACTGCCATGCATCACCATAGAGCTCCATTTGTTGGTCAGACATCG 516
 QY 196 GGTGAGGCGGAGCGCATTCGTCTGAAGGCACTGGCTGCAAGAGCGACGCTGATCCA 255
 Db 515 GGTGAGGCGGAGCGCATTCGTCTGAAGGCACTGGCTGCAAGAGCGACGCTGATCCA 456
 QY 256 CCACCGTGGCTGACATCTATCCAAACAGACTGCAATTTGAGTGCAGTGACCGAGGCAAG 315
 Db 455 CCACCGTGGCTGACATCTATCCAAACAGACTGCAATTTGAGTGCAGTGACCGAGGCAAG 396
 QY 316 CATGGATCCGCTCCGATGAAGATGATGGGGAGCGGGCTTCGTGGTCTGCTGCTC 375
 Db 395 CATGGATCCGCTCCGATGAAGATGATGGGGAGCGGGCTTCGTGGTCTGCTGCTC 336
 QY 376 CAACACCAAGCATGTTTCTGATTCGCTACGTATGAAAATTTGATCTACAGATATC 435
 Db 335 CAACACCAAGCATGTTTCTGATTCGCTACGTATGAAAATTTGATCTACAGATATC 276
 QY 436 CTGAGAGTGAGAGTTCATATTGAGTCTATTTTATCTTGTGATGTAATTGCTTTGCT 495
 Db 275 CTGAGAGTGAGAGTTCATATTGAGTCTATTTTATCTTGTGATGTAATTGCTTTGCT 216
 QY 496 TGTCCCTCAGAGTATTCGTTTGTGTTGGATGAGCAAGTGGAAATTAAGAGTGTACTA 555
 Db 215 TGTCCCTCAGAGTATTCGTTTGTGTTGGATGAGCAAGTGGAAATTAAGAGTGTACTA 156
 QY 556 TATACCATCATCTCTGTTTGAAGTTTGCAGTTCTGCAATTCATGTATCTGTAATTG 615
 Db 155 TATACCATCATCTCTGTTTGAAGTTTGCAGTTCTGCAATTCATGTATCTGTAATTG 96
 QY 616 ATGATGCTGGATTTCTATATTATCAATGTCATTATCTGTTGTAATAAAAAA 673
 Db 95 ATGATGCTGGATTTCTATATTATCAATGTCATTATCTGTTGTAATAAAAAA 38

RESULT 2

CB627453/c

LOCUS

777 bp mRNA linear EST 08-APR-2003

DEFINITION

OSIIEB02G20.r OSIIEB Oryza sativa (indica cultivar-group) cDNA
 clone OSIIEB02G20 3', mRNA sequence.

ACCESSION CB627453
 VERSION CB627453.1
 KEYWORDS GI:29622442
 SOURCE EST.

ORGANISM

Oryza sativa (indica cultivar-group)
 Oryza sativa (indica cultivar-group)
 Rukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzoideae; Oryza.

REFERENCE

1 (bases 1 to 777)
 Jantassuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Masur,B.,
 Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
 Large-scale identification of ESTs involved in the interaction
 between rice and Magnaporthe grisea
 Unpublished

COMMENT

Contact: Rod Wing
 Arizona Genomics Institute
 University of Arizona
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
 85721-0088, USA
 Tel: 520 626 3967
 Fax: 520 621 9288
 Email: http://genome.arizona.edu

PCR Primers

FORWARD: gta aaa cga cgg cca gtc
 BACKWARD: gga aac agc tat gac cat g
 Plate: 02 row: G column: 20
 Seq primer: gga aac agc tat gac cat g.

FEATURES

Location/Qualifiers
 1..777
 /organism="Oryza sativa (indica cultivar-group)"

/mol_type="mRNA"
 /cultivar="IR36"
 /db_xref="taxon:39946"
 /clone="OSIIEB02G20"
 /tissue_type="leaf"
 /dev_stage="3 week"
 /lab_host="DH10B"
 /clone_lib="OSIIEB"
 /note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:
 XhoI; 24 hrs after inoculation with Rice Blast (PO6-6.3)"

BASE COUNT

221 a 206 c 173 g 177 t

ORIGIN

Query Match 94.0%; Score 632.8; DB 14; Length 777;
 Best Local Similarity 98.8%; Pred. No. 2.7e-155;
 Matches 648; Conservative 0; Mismatches 7; Indels 1; Gaps 1;
 QY 15 TCGATGGCAGCAAC-TCCATCTTGGGTGTGGCTGAGCACACCAAGAAATTCCTCC 73
 Db 657 TGAATGGCAGCAACATCCCTCTTGGGTGTGGCTGAGCACACCAAGAAATTCCTCC 598
 QY 74 TCAGTGGTTTGCAGCTGTCCATGCCCTGTTCCATTCATTGGAATCTGAGGAAATCTGT 133
 Db 597 TCAGTGGTTTGCAGCTGTCCATGCCCTGTTCCATTCATTGGAATCTGAGGAAATCTGT 538
 QY 134 CAACATGCCCAAGACTGCCATGGCATTCACATAGAGCTCCATTTGTTGGTCAGACAAT 193
 Db 537 CAACATGCCCAAGACTGCCATGGCATTCACATAGAGCTCCATTTGTTGGTCAGACAAT 478
 QY 194 CGGTTGAGGGCGGAGCGCATTCGTCTGAAGGCACTGGCTGCAAGAGCGACCTGATTC 253
 Db 477 CGGTTGAGGGCGGAGCGCATTCGTCTGAAGGCACTGGCTGCAAGAGCGACCTGATTC 418
 QY 254 CACCACCGTGGCTGACATGTATCCAAACAAAGACTGCAAAATTCAGTGCACACCGGGCAA 313
 Db 417 CACCACCGTGGCTGACATGTATCCAAACAAAGACTGCAAAATTCAGTGCACACCGGGCAA 358
 QY 314 GGCATGGATCCGCTCGCATGAAGATGATGGCGGACCGGGCTTCGTGGTGGTCTCTGTC 373
 Db 357 GGCATGGATCCGCTCGCATGAAGATGATGGCGGACCGGGCTTCGTGGTGGTCTCTGTC 298

```

QY 374 TCCACCAACCAAGCATGTGTTTCTGATTCGTCACGTGATTCGAAATTTGATCTACCAGTA 433
DB 297 TCCACCAACCAAGCATGTGTTTCTGATTCGTCACGTGATTCGAAATTTGATCTACCAGTA 238
QY 434 TCCCTGAGAGTGGAGAGTTCGATATTCAGTCTATTTTATCTGTCGATGTAATTCGCTTGG 493
DB 237 TCCCTGAGAGTGGAGAGTTCGATATTCAGTCTATTTTATCTGTCGATGTAATTCGCTTGG 178
QY 494 CTTGTCCTCCCTCAGAGTATTCGTTTGTGTTGGATGAGACAGTGGAAATGAAGTGTCTAC 553
DB 177 CTTGTCCTCCCTCAGAGTATTCGTTTGTGTTGGATGAGACAGTGGAAATGAAGTGTCTAC 118
QY 554 TATATACAGCATATTCGTTTGTGTTAAAGTTTGCAGTTCCTCAGTTCATGTCGTAATTC 613
DB 117 TATATACAGCATATTCGTTTGTGTTAAAGTTTGCAGTTCCTCAGTTCATGTCGTAATTC 58
QY 614 TGATGATCTGGATTCCTACTATTTATCAATTCGTCATATCTGTCGTAATTTGTA 669
DB 57 TGATGATCTGGATTCCTACTATTTATCAATTCGTCATATCTGTCGTAATTTGTA 2

RESULT 3
AU068433
LOCUS AU068433 Rice callus Oryza sativa (japonica cultivar-group) cDNA
DEFINITION clone C30250_52, mRNA sequence.
ACCESSION AU068433.1 GI:5003284
VERSION 1
KEYWORDS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
REFERENCE 1 (bases 1 to 723)
AUTHORS Yamamoto, K. and Sasaki, T.
TITLES Rice cDNA from callus (1998)
JOURNAL Unpublished
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsasaki@abrr.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/
PROJECT "RGP"

FEATURES
source
1..723
organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="C30250_52"
/clone_lib="Rice callus"
/notes="Vector: pBluescript II SK+; Site 1: SalI; Site 2:
NotI; cDNA prepared from rice callus mRNAs by using
oligo(dT) as a primer and ligating to the SalI-NotI site
of pBluescript II SK+ phagemid."
BASE COUNT 183 a 154 c 183 g 203 t
ORIGIN
Query Match 93.7%; Score 630.6; DB 9; Length 723;
Best Local Similarity 99.2%; Pzed. No. 1e-154;
Matches 644; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 17 GATGCGCAGCAAC-TCCATCTCTGGGTTGGTGGCGTGGACACACAGAAATTCCTCCCTC 75
DB 43 GAATGCGCAGCAACATCCCTCTCTGGGTTGGTGGCGTGGACACACAGAAATTCCTCCCTC 102
QY 76 AGTGGTTTGAGTGTCCATGCGGCTGTTCCTCATTCATTCGATGTCGAGGAATTCGTGCA 135
DB 103 AGTGGTTTGAGTGTCCATGCGGCTGTTCCTCATTCATTCGATGTCGAGGAATTCGTGCA 162

```

```

QY 136 ACATGCCCAAGACTGCCATGGCATTCCATAGCAGCCTCCATTGTTGGTCAGCAATCG 195
DB 163 ACATGCCCAAGACTGCCATGGCATTCCATAGCAGCCTCCATTGTTGGTCAGCAATCG 222
QY 196 GGTCCGAGGGCGGAGCGCATTCGTTCTGAAGGCACATCGGCTCAAGAGCGACGCTGATTCCA 255
DB 223 GGTCCGAGGGCGGAGCGCATTCGTTCTGAAGGCACATCGGCTCAAGAGCGACGCTGATTCCA 282
QY 256 CCACCGTGGCTGCATGATTCCTCAAAACAGACACTGCAATTCGAGTGCACACGAGGGCAAGG 315
DB 283 CCACCGTGGCTGCATGATTCCTCAAAACAGACACTGCAATTCGAGTGCACACGAGGGCAAGG 342
QY 316 CATCGGATCCGCTCGCGATGAAGATGATCGCGGAGCGGCTTCGTGGTGGTGGTGGTTC 375
DB 343 CATCGGATCCGCTCGCGATGAAGATGATCGCGGAGCGGCTTCGTGGTGGTGGTGGTTC 402
QY 376 CAACCAACAGCATGTGTTCTGATTCCTGATTCCTGATTCCTGATTCCTGATTCCTGATTC 435
DB 403 CAACCAACAGCATGTGTTCTGATTCCTGATTCCTGATTCCTGATTCCTGATTCCTGATTC 462
QY 436 CCTCGAGAGTGGAGAGTTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 495
DB 463 CCTCGAGAGTGGAGAGTTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 522
QY 496 TGTCCCTCAGAGTATTCGTTGTTGGGATGAGACAGTGCATTAAGAGTGCCTACTA 555
DB 523 TGTCCCTCAGAGTATTCGTTGTTGGGATGAGACAGTGCATTAAGAGTGCCTACTA 582
QY 556 TATACACGATCATTCCTGTTGAAGTTTGCAGTTCGAGTTCATGTCATGTCATGTCATGTCAT 615
DB 583 TATACACGATCATTCCTGTTGAAGTTTGCAGTTCGAGTTCATGTCATGTCATGTCATGTCAT 642
QY 616 ATGATGCTGGATTCCTACTATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 664
DB 643 ATGATGCTGGATTCCTACTATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 691

RESULT 4
CB000367
LOCUS CB000367 776 bp mRNA linear EST 10-JAN-2003
DEFINITION S345U_H05 Rice cold stress germination cDNA library Oryza sativa
(japonica cultivar-group) cDNA 5', mRNA sequence.
ACCESSION CB000367.1 GI:27577672
VERSION 1
KEYWORDS Oryza sativa (japonica cultivar-group)
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
REFERENCE 1 (bases 1 to 776)
AUTHORS de los Reyes, B.G., Morsy, M., Gibbons, J., Varma, T.S.N., Antoine, M.,
Redus, M., McGrath, J.M. and Halgren, R.
TITLES Development of a chilling stress EST library of germinating rice
(Oryza sativa L. subsp. japonica) enriched with stress-related and
novel genes
JOURNAL Unpublished
COMMENT Contact: Benildo G. de los Reyes
Plant Genomics Lab., Department of Crop, Soil and Environmental
Sciences
University of Arkansas
115 Plant Science Building, Fayetteville, AR 72701, USA
Tel: (479)-575-7465
Fax: (479)-575-8435
Email: breyes@uark.edu
Plate: S345U row: H column: 05
Seq primer: 73
Location/Qualifiers
1..776
organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="P1560247"
/db_xref="taxon:39947"

```

```

/issue_type="coleotile, radicle, prophyll, immature leaf"
/dev_stage="embryo at different stages of germination and
shoots of germinated seeds under chilling stress
(130C/100C)"
/lab_host="SOLR"
/clone_lib="Rice cold stress germination cDNA library"
/notes="Organ: seedlings; Vector: Lambda Uni-Zap XR excised
as pBluescript; Site 1: EORI; Site 2: XhoI; The cDNAs
were derived from reverse transcription of mRNA samples
from seeds at different stages of germination and
seedlings at early phase of growth under chilling stress
(130C/100C). The mRNA pool was used as template for double
stranded cDNA synthesis using the Stratagene Uni-Zap XR
cDNA synthesis and library kit. A total of 150,000 phages
were excised from the primary library as pBluescript
phagenid clones. Enrichment of the primary excised library
with chilling-induced transcripts was performed by
hybridizing the primary excised library colony lifts with
the PCR-select subtraction product, with cold germinated
cDNA as tester and control temperature-germinated cDNA as
driver."
BASE COUNT      170 a   180 c   203 g   223 t
ORIGIN
Query Match      93.7%; Score 630.6; DB 14; Length 776;
Best Local Similarity 99.2%; Pred. No. 1e-154;
Matches 644; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

```

```

QY 17 GATGGAGCAAGC-TCCATCTTGGGGTGGGGTGGAGCACACCAAGAAATTCCTCCCTC 75
DB 118 GAATGGCAGCAACATCCCTCTTGGGGTGGGGTGGAGCACACCAAGAAATTCCTCCCTC 177
QY 76 AGTGGTTTGCAGCTGCCATCGCGCTGTTCCATTCATTGGAATGCTGAGAAATCTGTCA 135
DB 178 AGTGGTTTGCAGCTGCCATCGCGCTGTTCCATTCATTGGAATGCTGAGAAATCTGTCA 237
QY 136 ACATGCCAAGACTGCCATGCGATTCACCATAGCAGCTCCATTTGTGTGACAAATCG 195
DB 238 ACATGCCAAGACTGCCATGCGATTCACCATAGCAGCTCCATTTGTGTGACAAATCG 297
QY 196 GPTCGAGGCGGAGCGCATTCGCTGAAGGACTGCTGCAAGAGCGAGCTGATTCCA 255
DB 298 GGTCTGAGGCGGAGCGCATTCGCTGAGGAGCTGCTGCAAGAGCGAGCTGATTCCA 357
QY 256 CCACCGTGGTGCATGTATCCAAACAAGACTGCAAAATTCAGTGTACACCGAGGCAAGG 315
DB 358 CCACCGTGGTGCATGTATCCAAACAAGACTGCAAAATTCAGTGTACACCGAGGCAAGG 417
QY 316 CATGGGATCCGCTCGGATGAGATGATGCGGAGCGGGCTTCTGTGTGTGCTGTGCTC 375
DB 418 CATGGGATCCGCTCGGATGAGATGATGCGGAGCGGGCTTCTGTGTGTGCTGTGCTC 477
QY 376 CAACACCAAGCATGTGTTCTGATTCCTACTGATGGAAATTCGTATCTACCAATATC 435
DB 478 CAACACCAAGCATGTGTTCTGATTCCTACTGATGGAAATTCGTATCTACCAATATC 537
QY 436 CTTGGAGAGTGGAGAGTTGATATTTGATCTATTTTATCTTGATCTAAATTCCTTTGCT 495
DB 538 CTTGGAGAGTGGAGAGTTGATATTTGATCTATTTTATCTTGATCTAAATTCCTTTGCT 597
QY 496 TGTCCCTCAGAGATATTCGTTTGTGGATGAGACCAAGTGGAAATTAAGAGTGTACTA 555
DB 598 TGTCCCTCAGAGATATTCGTTTGTGGATGAGACCAAGTGGAAATTAAGAGTGTACTA 657
QY 556 TATACAGCATCTCTGTTTGAAGTTTGCAGTCTTCTGAGTTTCATGATCTGTATTTTG 615
DB 658 TATACAGCATCTCTGTTTGAAGTTTGCAGTCTTCTGAGTTTCATGATCTGTATTTTG 717
QY 616 ATGATGCTGGATTTCTACTATTTATCAATCGTCATTTACTGTGTGTAA 664
DB 718 ATGATGCTGGATTTCTACTATTTATCAATCGTCATTTACTGTGTGTAA 766

```

```

RESULT 5
AUI65747
LOCUS      AUI65747       705 bp      mRNA      linear      EST 03-APR-2002
DEFINITION      AUI65747 Rice panicle at flowering stage Oryza sativa (japonica
                  cultivar-group) cDNA clone B4287, mRNA sequence.
ACCESSION      AUI65747.1 GI:11565111
VERSION        AUI65747
KEYWORDS       EST.
SOURCE         Oryza sativa (japonica cultivar-group)
ORGANISM       Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE      1 (bases 1 to 705)
AUTHORS        Sasaki, T. and Yamamoto, K.
TITLE          Rice cDNA from panicle at flowering stage (2000)
JOURNAL        Unpublished
COMMENT        Contact: Takuji Sasaki
                  National Institute of Agrobiological Resources
                  Rice Genome Research Program, Kamondai 2-1-2, Tsukuba, Ibaraki
                  305-8602, Japan
                  Tel: 81-298-38-7441
                  Fax: 81-298-38-7468
                  Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
                  PROJECT = "RGP".
FEATURES             Location/Qualifiers
                     1..705
                        /organism="Oryza sativa (japonica cultivar-group)"
                        /molecule="mRNA"
                        /cultivar="Nipponbare"
                        /db_xref="taxon:39947"
                        /clone="E4287"
                        /dev_stage="flowering stage"
                        /clone_lib="Rice panicle at flowering stage"
                        /note="Organ: panicle; Rice cDNA from panicle at flowering
                        stage"
BASE COUNT      175 a   150 c   178 g   199 t   3 others
ORIGIN
Query Match      93.4%; Score 628.6; DB 9; Length 705;
Best Local Similarity 99.1%; Pred. No. 3.4e-154;
Matches 642; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

```

```

QY 27 AAATCCATCTTGGGGTGTGGCGTGGAGCACACCAAGAAATTCCTCCCTCAGTGGTTGCA 86
DB 52 AACATCCNTCTTGGGGTGTGGCGTGGAGCACACCAAGAAATTCCTCCCTCAGTGGTTGCA 111
QY 87 GCTGTCCATGCGCTGTTT-CCATTCATTGGGAATGCTGAGGAATCTGTCAACATGCCCAA 145
DB 112 GCTGTCCATGCGCTGTTTCCATTCATTGGGAATGCTGAGGAATCTGTCAACATGCCCAA 171
QY 146 GACTGCCATGGCATTCACCATAGAGCTCTCATTTGTTGGTCAGACAATTCGGGTTCAGGGC 205
DB 172 GACTGCCATGGCATTCACCATAGAGCTCTCATTTGTTGGTCAGACAATTCGGGTTCAGGGC 231
QY 206 GGAGGGCATTCGTCTGAAGGCACTGGCTGCAAGAGCGAGCTCATTCACACCGTGGC 265
DB 232 GGAGGGCATTCGTCTGAAGGCACTGGCTGCAAGAGCGAGCTCATTCACACCGTGGC 291
QY 266 TGACATGATCCAAACAAGACTGCAAAATTCAGTGCACACCGAGGGCAAGCATGGATCC 325
DB 292 TGACATGATCCAAACAAGACTGCAAAATTCAGTGCACACCGAGGGCAAGCATGGATCC 351
QY 326 CCTCGCATGAAGATGATGGCGGACGGGCTTCTGGTGGTCTGCTCTCCCAACACCAAG 385
DB 352 GCTCGCATGAAGATGATGGCGGACGGGCTTCTGGTGGTCTGCTCTCCCAACACCAAG 411
QY 386 CATGTGTTTCTGATGCTCACTGATTTGGAAATTTGTATCTTACCAGTATCCCTGGAGGT 445
DB 412 CATGTGTTTCTGATGCTCACTGATTTGGAAATTTGTATCTTACCAGTATCCCTGGAGGT 471
QY 446 GGAGAGTTGATTTGAGTCTATTTTATCTTGTGATTAATTCCTTTGCTGCTCCCTCAG 505

```


255	ACCACCGTGGCTGACATGTATCCAAACAAGACTGCANAATTGCAGTGACACCGAGGCCAAG	314	
Ddb	289	ACCACCGTGGCTGACATGTATCCAAACAAGACTGCANAATTGCAGTGACACCGAGGCCAAG	348
QY	315	GCATGGGATCCGCTCCGATGAAGATGATGGCGGACGGGCTTCCTGGTGGTGGTGGTGGTGGT	374
Ddb	349	GCATGGGATCCGCTCCGATGAAGATGATGGCGGACGGGCTTCCTGGTGGTGGTGGTGGTGGT	408
QY	375	CAACACCAAGCATGTGTTCTGATTCCTCAGTGAATGGAAAATTTGTATCTACCAAGTAT	434
Ddb	409	CAACACCAAGCATGTGTTCTGATTCCTCAGTGAATGGAAAATTTGTATCTACCAAGTAT	468
QY	435	CCCTGGAGAGTGGAGAGTTGATTAATGAGTCTATTTTATCTCTGATGATAATTGCCCTTTC	494
Ddb	469	CCCTGGAGAGTGGAGAGTTGATTAATGAGTCTATTTTATCTCTGATGATAATTGCCCTTTC	528
QY	495	TGTCCCTCAGAAGTATTCGTTGTTGTTGGGATGAGCAAGTGAATGAAGAGTGTCTACT	554
Ddb	529	TGTCCCTCAGAAGTATTCGTTGTTGTTGGGATGAGCAAGTGAATGAAGAGTGTCTACT	588
QY	555	ATATACAGATCATTCCTGTTGTTAAGTTTGGCAGTCTTGCAGTTCATGATCTCTGTAATTT	614
Ddb	589	ATATACAGATCATTCCTGTTGTTAAGTTTGGCAGTCTTGCAGTTCATGATCTCTGTAATTT	648
QY	615	GATGATGCTGATTTCTACTATTTATCAATGCTCATTTATCTGTTGTTAAAAA	673
Ddb	649	GATGATGCTGATTTCTACTATTTATCAATGCTCATTTATCTGTTGTTAAAAA	707

RESULT 7
CB660999/c

LOCUS CB660999.1 805 bp mRNA linear EST 09-APR-2003

DEFINITION OSUNED03E21.r OSUNED Oryza sativa (japonica cultivar-group) cDNA clone OSUNED03E21.3, mRNA sequence.

ACCESSION CB660999

KEYWORDS CB660999.1 GI:29664724

SOURCE EST.

ORGANISM Oryza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.

1 (bases 1 to 805)

Jatnasuriyar,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E., Kundra,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.

Large-scale identification of ESTs involved in the interaction between rice and Magnaporthe grisea

Unpublished

Contact: Rod Wing

Arizona Genomics Institute

University of Arizona

Biological Sciences West, 448H, P.O. Box 210088, Tucson, AZ 85721-0088, USA

Tel.: 520 626 3967

Fax: 520 621 9288

Email: http://genome.arizona.edu

PCR Primers

FORWARD: gta aaa cga cgg cca.gtg

REVERSE: gga aac agc tat gac cat g

Plate: 03 row: B column: 21

Seq primer: gga aac agc tat gac cat g.

Location/Qualifiers

1..805

/organism="Oryza sativa (japonica cultivar-group)"

/mol_type="mRNA"

/cultivar="Nipponbare"

/db_xref="taxon:39947"

/clone="OSUNED03E21"

/tissue_type="Leaf"

/dev_stage="3 week"

/lab_host="DH10B"

/clone_lib="DS10B"

/library="EST"

source

/note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2: XhoI; 24 hrs after inoculation with Rice Blast (C9240-1)"

226 a 212 c 190 g 177 t

Query Match 93.4%; Score 628.6; DB 14; Length 805;
Best Local Similarity 99.2%; Pred. No. 3.5e-154;
Matches 642; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 17 GGTGTCAGCAAC-TCCATCTTGGGGTGGGGTGGAGCAGCAGCAAGAAATTCCTCCCTC 75
DB 647 GAATGTCAGCAACATCCCTCTTGGGGTGGGGTGGAGCAGCAGCAAGAAATTCCTCCCTC 588
QY 76 AGTGGTTTGAGCTGTCATCCGCTCTTCCATTCATTTGGAAATGCTGAGAAATCTGTCA 135
DB 587 AGTGGTTTGAGCTGTCATCCGCTCTTCCATTCATTTGGAAATGCTGAGAAATCTGTCA 528
QY 136 ACATGCCAGAGCTGCATGCAATCCATTCAGCAGCTCCATTTGGTTCAGCAATCG 195
DB 527 ACATGCCAGAGCTGCATGCAATCCATTCAGCAGCTCCATTTGGTTCAGCAATCG 468
QY 196 GGTGTCAGCAAC-TCCATCTTGGGGTGGGGTGGAGCAGCAGCAAGAAATTCCTCCCTC 255
DB 467 GAATGTCAGCAACATCCCTCTTGGGGTGGGGTGGAGCAGCAGCAAGAAATTCCTCCCTC 408
QY 256 CCACCGTGGCTGACATGATATCCAAACAGAGCTGCAATTCAGTGCAGCAGGCGAAG 315
DB 407 CCACCGTGGCTGACATGATATCCAAACAGAGCTGCAATTCAGTGCAGCAGGCGAAG 348
QY 316 CATGGGATCCGCTCGCATGAAGATGATGGGAGCGGCTTCTGGTGGTCTGCTGCTC 375
DB 347 CATGGGATCCGCTCGCATGAAGATGATGGGAGCGGCTTCTGGTGGTCTGCTGCTC 288
QY 376 CAACACCAAGCATGTTCTGATGCTCACTGATTCGAAATTTGATCTACAGTATC 435
DB 287 CAACACCAAGCATGTTCTGATGCTCACTGATTCGAAATTTGATCTACAGTATC 228
QY 436 CTTGGAGAGTGGAGAGTTGATTTAGTCTATTTTATCTTGTGATGATTTGCTTTGCT 495
DB 227 CTTGGAGAGTGGAGAGTTGATTTAGTCTATTTTATCTTGTGATGATTTGCTTTGCT 168
QY 496 TGTCCCTCAGAAATTCGTTTGGTGGTGGAGCAAGAGTGAAGTGAAGTGTCTACTA 555
DB 167 TGTCCCTCAGAAATTCGTTTGGTGGTGGAGCAAGAGTGAAGTGAAGTGTCTACTA 108
QY 556 TATACAGATCATCTGTTTGAAGTTGGCAGTTCGAGTTCATGATCTGTAATTG 615
DB 107 TATACAGATCATCTGTTTGAAGTTGGCAGTTCGAGTTCATGATCTGTAATTG 48
QY 616 ATGATGCTGGAATTCATATTTATCAATGCTCATTTATCTGTTGTT 662
DB 47 ATGATGCTGGAATTCATATTTATCAATGCTCATTTATCTGTTGTT 1

RESULT 8
CB647623/c
LOCUS CB647623 682 bp mRNA linear EST 08-APR-2003
DEFINITION OSJNEB10K01.r OSJNEB Oryza sativa (japonica cultivar-group) cDNA
clone OSJNEB10K01.3, mRNA sequence.
ACCESSION CB647623
VERSION CB647623.1 GI:29642616
KEYWORDS EST.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 682)
Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
Kudrna,D., Dean,R., Soderlund,C., Wang,R. and Wang,G.
TITLE Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe oryzae
JOURNAL Unpublished

COMMENT

Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR Primers
FORWARD: gta aaa cga cgg cca gtc
BACKWARD: gga aac agc tat gac cat g
Plate: 10 row: K column: 01
Seq primer: gga aac agc tat gac cat g.
Location/Qualifiers
1. 682
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSJNEB10K01"
/tissue_type="Leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="OSJNEB"
/note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:
XhoI; 24 hrs after inoculation with Rice Blast (Che 86061
)*"

BASE COUNT 199 a 179 c 145 g 159 t

Query Match 93.2%; Score 627.4; DB 14; Length 682;
Best Local Similarity 98.9%; Pred. No. 7e-154;
Matches 642; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

QY 17 GGTGTCAGCAAC-TCCATCTTGGGGTGGGGTGGAGCAGCAGCAAGAAATTCCTCCCTC 75
DB 653 GAATGTCAGCAACATCCCTCTTGGGGTGGGGTGGAGCAGCAGCAAGAAATTCCTCCCTC 594
QY 76 AGTGGTTTGAGCTGTCATCCGCTCTTCCATTCATTTGGAAATGCTGAGAAATCTGTCA 135
DB 593 AGTGGTTTGAGCTGTCATCCGCTCTTCCATTCATTTGGAAATGCTGAGAAATCTGTCA 534
QY 136 ACATGCCAGAGCTGCATGCGATTCACATAGAGCTCCATTTGGTTCAGCAATCG 195
DB 533 ACATGCCAGAGCTGCATGCGATTCACATAGAGCTCCATTTGGTTCAGCAATCG 474
QY 196 GGTGTCAGCAAC-TCCATCTTGGGGTGGGGTGGAGCAGCAGCAAGAAATTCCTCCCTC 255
DB 473 GGTGTCAGCAAC-TCCATCTTGGGGTGGGGTGGAGCAGCAGCAAGAAATTCCTCCCTC 414
QY 256 CCACCGTGGCTGACATGATATCCAAACAGAGCTGCAATTTGGTTCAGCAATCG 315
DB 413 CCACCGTGGCTGACATGATATCCAAACAGAGCTGCAATTTGGTTCAGCAATCG 354
QY 316 CATGGGATCCGCTCGCATGAAGATGATGGGAGCGGCTTCTGGTGGTCTGCTGCTC 375
DB 353 CATGGGATCCGCTCGCATGAAGATGATGGGAGCGGCTTCTGGTGGTCTGCTGCTC 294
QY 376 CAACACCAAGCATGTTTCTGATTCATCTGATTTGGAAATTTGATCTACAGTATC 435
DB 293 CAACACCAAGCATGTTTCTGATTCATCTGATTTGGAAATTTGATCTACAGTATC 234
QY 436 CTTGGAGAGTGGAGAGTTGATTTAGTCTATTTTATCTTGTGATGATTTGCTTTGCT 495
DB 233 CTTGGAGAGTGGAGAGTTGATTTAGTCTATTTTATCTTGTGATGATTTGCTTTGCT 174
QY 496 TGTCCCTCAGAAATTCGTTTGGTGGGATGAGCAAGTGAAGTGAAGTGTCTACTA 555
DB 173 TGTCCCTCAGAAATTCGTTTGGTGGGATGAGCAAGTGAAGTGAAGTGTCTACTA 114
QY 556 TATACAGATCATCTGTTTGAAGTTGGCAGTTCGAGTTCATGATCTGTAATTG 615
DB 113 TATACAGATCATCTGTTTGAAGTTGGCAGTTCGAGTTCATGATCTGTAATTG 54

QY 616 ATGATGCTGGATTCTTACTATTATCAATGTCATTTATCTGTA 664
 Db 53 ATGATGCTGGATTCTTACTATTATCAATGTCATTTATCTGTA 5

RESULT 9
 C8663774/c
 LOCUS
 DEFINITION OSJNE09B20.r OSJNE09B20 3', mRNA sequence.
 ACCESSION C8663774.1 GI:29667499
 VERSION
 KEYWORDS EST.
 SOURCE
 ORGANISM
 Oriza sativa (japonica cultivar-group)
 Oriza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzeae; Oriza.
 1 (bases 1 to 763)
 Jantassuriyarat, C., Lu, G., Gowda, M., Hatfield, J., Zhou, B., Mazur, E.,
 Kudrna, D., Dean, R., Soderlund, C., Wang, R. and Wang, G.
 Large-scale identification of ESTs involved in the interaction
 between rice and Magnaporthe grisea
 Unpublished
 CONTACT: Rod Wing
 ARIZONA: Genomics Institute
 UNIVERSITY OF ARIZONA
 BIOLOGICAL SCIENCES WEST, 448A, P.O. Box 210088, Tucson, AZ
 85721-0088, USA
 Tel: 520 626 3967
 Fax: 520 621 9288
 Email: http://genome.arizona.edu
 PCR PRIMERS
 FORWARD: gta aaa cga cga cca gtc
 BACKWARD: gga aac agc tat gac cat g
 Plate: 09 row: B column: 20
 Seq primer: gga aac agc tat gac cat g.
 Location/Qualifiers
 1..763
 /organism="Oriza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="OSJNE09B20"
 /tissue_type="Leaf"
 /dev_stage="3 week"
 /lab_host="DH108"
 /clone_lib="OSJNEd"
 /note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:
 XhoI; 24 hrs after inoculation with Rice Blast (Cs240-1)"
 XhoI 217 a 202 c 177 g 167 t

BASE COUNT 217 a 202 c 177 g 167 t

ORIGIN
 Query Match 93.2%; Score 627; DB 14; Length 763;
 Best Local Similarity 99.1%; Pred. No. 9,1e-154;
 Matches 641; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 17 GGATGGCAGCAAC-TCCATCTTGGGGTGTGGCGTGAGCACACCAAGAAATTCCTCCCTC 75
 Db 647 GAATGGCAGCAACATCCCTCTTGGGGTGTGGCGTGAGCACACCAAGAAATTCCTCCCTC 588

QY 76 AGTGGTTGAGCTGTCATGGCGCTGTCATTCATTCGATGCTGAGGAATCTGCA 135
 Db 587 AGTGGTTGAGCTGTCATGGCGCTGTCATTCATTCGATGCTGAGGAATCTGCA 528

QY 136 ACATGCCCAAGACTGCCATGGCATTCCATAGCAGCTCCATTTGTTGTCAGACAATCG 195
 Db 527 ACATGCCCAAGACTGCCATGGCATTCCATAGCAGCTCCATTTGTTGTCAGACAATCG 468

QY 196 GGTGCGAGGGCGGAGCGGCTGTCAGAGGCTGCTGCAAGGCGGCTGCAAGGCGGCTGCA 255
 Db 467 GGTGCGAGGGCGGAGCGGCTGTCAGAGGCTGCTGCAAGGCGGCTGCAAGGCGGCTGCA 408

QY 256 CCACCGTGGCTGACATGTATCCAAACAAGACTGCAAAATTCAGTGCACCCGAGGGCAAGG 315
 Db 407 CCCCCTGGCTGACATGTATCCAAACAAGACTGCAAAATTCAGTGCACCCGAGGGCAAGG 348

QY 316 CATGGATCCGCTCGCATGAAGATGATGGCGGAGCGGCTGCTGGTGGTGGTGGTGGTGG 375
 Db 347 CATGGATCCGCTCGCATGAAGATGATGGCGGAGCGGCTTCTGGTGGTGGTGGTGGTGG 288

QY 376 CAACACCAAGCATGTGTTTCTGATTTCTCACTCAATTTGGAAATTTGTATCTACAGTATC 435
 Db 287 CAACACCAAGCATGTGTTTCTGATTTCTCACTCAATTTGGAAATTTGTATCTACAGTATC 228

QY 436 CTTGGAGATGGAGATGATGATTTGATGCTATTTATCTTGTGATGATGATGATGATGATG 495
 Db 227 CTTGGAGATGGAGATGATGATTTGATGCTATTTATCTTGTGATGATGATGATGATGATG 168

QY 496 TGTCCCTCAGAAATTTCTGTTTGTGGATGAGCAAGTGGAAATTAAGAGTGTCTACTA 555
 Db 167 TGTCCCTCAGAAATTTCTGTTTGTGGATGAGCAAGTGGAAATTAAGAGTGTCTACTA 108

QY 556 TATACAGATCATTTCTGTTTGTAAATTTGTCAGTTTCTGCAATTCATGATCTGTAATTTG 615
 Db 107 TATACAGATCATTTCTGTTTGTAAATTTGTCAGTTTCTGCAATTCATGATCTGTAATTTG 48

QY 616 ATGATGCTGGATTTCTACTATTATCAATGCTCATTTATCTGTTGT 662
 Db 47 ATGATGCTGGATTTCTACTATTATCAATGCTCATTTATCTGTTGT 1

RESULT 10
 C8649882/c
 LOCUS
 DEFINITION OSJNEB14B02.r OSJNEB14B02 3', mRNA sequence.
 ACCESSION C8649882
 VERSION
 KEYWORDS EST.
 SOURCE
 ORGANISM
 Oriza sativa (japonica cultivar-group)
 Oriza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzeae; Oriza.
 1 (bases 1 to 768)
 Jantassuriyarat, C., Lu, G., Gowda, M., Hatfield, J., Zhou, B., Mazur, E.,
 Kudrna, D., Dean, R., Soderlund, C., Wang, R. and Wang, G.
 Large-scale identification of ESTs involved in the interaction
 between rice and Magnaporthe grisea
 Unpublished
 CONTACT: Rod Wing
 ARIZONA: Genomics Institute
 UNIVERSITY OF ARIZONA
 BIOLOGICAL SCIENCES WEST, 448A, P.O. Box 210088, Tucson, AZ
 85721-0088, USA
 Tel: 520 626 3967
 Fax: 520 621 9288
 Email: http://genome.arizona.edu
 PCR PRIMERS
 FORWARD: gta aaa cga cga cca gtc
 BACKWARD: gga aac agc tat gac cat g
 Plate: 14 row: B column: 02
 Seq primer: gga aac agc tat gac cat g.
 Location/Qualifiers
 1..768
 /organism="Oriza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="OSJNEB14B02"
 /tissue_type="Leaf"
 /dev_stage="3 week"
 /lab_host="DH108"
 /clone_lib="OSJNEd"

FEATURES
 source
 1..768
 /organism="Oriza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="OSJNEB14B02"
 /tissue_type="Leaf"
 /dev_stage="3 week"
 /lab_host="DH108"
 /clone_lib="OSJNEd"

```

/note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:
XhoI; 24 hrs after inoculation with Rice Blast (Che 86061
)"
BASE COUNT      229 a   202 c   157 g   180 t
ORIGIN
Query Match      92.0%; Score 619.4; DB 14; Length 768;
Best Local Similarity 98.2%; Pred. No. 9.1e-152;
Matches 637; Conservative 0; Mismatches 11; Indels 1; Gaps 1;
QY 17 GGATGGCAGCAAAAC-TCCATCTTGGGCTGGCGTGAGCACACCAAGAAATTCCTCCCTC 75
DB 740 GAATGGCAGCAAAACATCCCTCTTGGGCGTGCGTGAGCACACCAAGAAATTCCTCCCTC 681
QY 76 AGTGGTTTCAGCTGCCATGGCTGTTCCATTCATTCGATTCGATTCGAGGAATCTGTCA 135
DB 680 AGTGGTTTCAGCTGCCATGGCTGTTCCATTCATTCGATTCGAGGAATCTGTCA 621
QY 136 ACATGCCCAAGACTGCCATGGCAATCCACATAGACGCTCCCATTTGTTGTCAGACAATCG 195
DB 620 ACATGCCCAAGACTGCCATGGCAATCCACATAGACGCTCCCATTTGTTGTCAGACAATCG 561
QY 196 GGTGGAGGGCGGAGCGCATCTGCTCAAGGCACTGCTGCAAGGCGGAGCGCTGATTCGA 255
DB 560 GGTGGAGGGCGGAGCGCATCTGCTGCAAGGCACTGCTGCAAGGCGGAGCGCTGATTCGA 501
QY 256 CCACCGTGGCTGACATGTATCCAAACAAGACTGCAAAATTCGAGTGACACCGAGGCAAGG 315
DB 500 CCACCGTGGCTGACATGTATCCAAACAAGACTGCAAAATTCGAGTGACACCGAGGCAAGG 441
QY 316 CATGGATCGCTCGGATGAGATGATGCGGAGCGGCTTCTGGTGCTGCTGCTC 375
DB 440 CATGGATCGCTCGGATGAGATGATGCGGAGCGGCTTCTGGTGCTGCTGCTC 381
QY 376 CAACACCAAGACTGCTGTTCTGATTCCTCACTGATTGGAATTTGTATCTACCAAGTATC 435
DB 380 CAACACCAAGACTGCTGTTCTGATTCCTCACTGATTGGAATTTGTATCTACCAAGTATC 321
QY 436 CTGGAGAGTGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 495
DB 320 CTGGAGAGTGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 261
QY 496 TGTCCCTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 555
DB 260 TGTCCCTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 201
QY 556 TATACAGATCATCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 615
DB 200 TATACAGATCATCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 141
QY 616 ATGATGCTGATTTCTACTATTATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 664
DB 140 ATGATGCTGATTTCTACTATTATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 92

RESULT 11
CB647260/c
LOCUS
DEFINITION
OSJNEb10821.r OSJNEb Oryza sativa (japonica cultivar-group) cDNA
clone OSJNEb10821.3', mRNA sequence.
ACCESSION
CB647260
VERSION
CB647260.1
KEYWORDS
EST.
SOURCE
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE
1 (bases 1 to 792)
Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea

```

```

JOURNAL
COMMENT
Unpublished
Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR Primers
FORWARD: gta aac cga cgg cca gtg
BACKWARD: gga aac agc tat gac cat g
Plate: 10 row: B column: 21
Seq primer: gga aac agc tat gac cat g.
Location/Qualifiers
1..792
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSJNEb10821"
/tissue_type="Leaf"
/dev_stage="3 week"
/lab_hosts="DH10B"
/clone_lib="OSJNEb"
/note="vector: pBluescript II KS +; Site 1: EcoRI; Site 2:
XhoI; 24 hrs after inoculation with Rice Blast (Che 86061
)"
BASE COUNT      222 a   212 c   184 g   174 t
ORIGIN
Query Match      91.3%; Score 614.4; DB 14; Length 792;
Best Local Similarity 98.8%; Pred. No. 1.9e-150;
Matches 640; Conservative 0; Mismatches 6; Indels 2; Gaps 2;
QY 17 GGATGGCAGCAAAAC-TCCATCTTGGGCTGGCGTGAGCACACCAAGAAATTCCTCCCTC 75
DB 648 GAATGGCAGCAAAACATCCCTCTTGGGCTGGCGTGAGCACACCAAGAAATTCCTCCCTC 589
QY 76 AGTGGTTTCAGCTGCTCCATGCGCTGTTCCATTCATTCGAAATGCTGAGGAATCTGTCA 135
DB 588 AGTGGTTTCAGCTGCTCCATGCGCTGTTCCATTCATTCGAAATGCTGAGGAATCTGTCA 529
QY 136 ACATGCCCAAGACTGCCATGCTACCATAGCAGCTCCATTTGTTGGTCAGACAATCG 195
DB 528 ACATGCCCAAGACTGCCATGCTACCATAGCAGCTCCATTTGTTGGTCAGACAATCG 469
QY 196 GGTGAGGGCGGAGCGCATTCGTTCTGAAGGCACTGGCTGCAAGAGGCACTGATTCCTCA 255
DB 468 GGTGAGGGCGGAGCGCATTCGTTCTGAAGGCACTGGCTGCAAGAGGCACTGATTCCTCA 409
QY 256 CCACCGTGGCTGACATGTATCCAAACAAGACTGCAAAATTCGAGTGACACCGAGGCAAGG 315
DB 408 CCACCGTGGCTGACATGTATCCAAACAAGACTGCAAAATTCGAGTGACACCGAGGCAAGG 349
QY 316 CATGGATCGCTCGCATGAAGATGATGCGGGAACGGCTTCTGGTGCTGCTGCTC 375
DB 348 CATGGATCGCTCGCATGAAGATGATGCGGGAACGGCTTCTGGTGCTGCTGCTC 289
QY 376 CAACACCAAGCATGTTGTTCTGATTCGCTCACTGATTCGGAATTTGTATCTACCAAGTATC 435
DB 288 CAACACCAAGCATGTTGTTCTGATTCGCTCACTGATTCGGAATTTGTATCTACCAAGTATC 229
QY 436 CCTGGAGAGTGGAGATTTGATATTGATCTATTATTTCTTGTGATGATTAATTTGCTTTGCT 495
DB 228 CCTGGAGAGTGGAGATTTGATATTGATCTATTATTTCTTGTGATGATTAATTTGCTTTGCT 169
QY 496 TGTCCCTCAGAGATTTGCTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 555
DB 168 TGTCCCTCAGAGATTTGCTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 109
QY 556 TATACAGATCATCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 615

```

```

Db      108  TGTACAGATCATCTCTGTTAGTTGGAGTTTCCAGTTCTGCGATTCAATGATCTCTAATTG 49
QY      616  ATGATGCTGATTTC-TACTATTATCAATGCTCATTATATCTGTTGT 662
Db      48  ATGATGCTGATTTCGTTGCTATTATCAATGCTCATTATATCTGTTGT 1
RESULT 12
LOCUS   CB655085/c
DEFINITION OSJNEC08C12.r OSJNEC Oryza sativa (japonica cultivar-group) cDNA
ACCESSION CB655085
VERSION   CB655085.1 GI:29658810
KEYWORDS  EST.
SOURCE    Oryza sativa (japonica cultivar-group)
ORGANISM  Oryza sativa (japonica cultivar-group)
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Eriartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 727)
AUTHORS   Jantaeuivat.C., Lu.G., Gowda.M., Hatfield.J., Zhou.B., Mazur.B.,
Kudrna.D., Dean.R., Soderlund.C., Wing.R. and Wang.G.
TITLE     Large-scale identification of ESTs involved in the interaction
          between rice and Magnaporthe grisea
JOURNAL   Unpublished
COMMENT   Contact: Rod Wing
          Arizona Genomics Institute
          University of Arizona
          Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
          85721-0088, USA
          Tel: 520 626 3967
          Fax: 520 621 9288
          Email: http://genome.arizona.edu
PCR PRIMERS
FORWARD: gta aac cga cgg cca gtc
BACKWARD: gga aac agc tat gac cat g
Plate: 08 row: C column: 12
Seq primer: gga aac agc tat gac cat g.
FEATURES
    source
        1..727
            /organism="Oryza sativa (japonica cultivar-group)"
            /mol_type="mRNA"
            /cultivar="Nipponbare"
            /db_xref="taxon:39947"
            /clone="OSJNEC08C12"
            /tissue_type="Leaf"
            /dev_stage="3 week"
            /lab_hosts="DH10B"
            /note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:
            XhoI; 6 hrs after inoculation with Rice-Blast (C9240-1)"
BASE COUNT 204 a 192 c 166 g 165 t
Query Match 91.2%; Score 613.8; DB 14; Length 727;
Best Local Similarity 98.7%; Pred. No. 2,7e-150;
Matches 629; Conservative 0; Mismatches 7; Indels 1; Gaps 1;
QY      17  GGATGGCAGCAAC-TCCATCTTGGGCTGTGGCTGAGCACACCAAGAAATTCCTCCCTC 75
Db      637  GAATGGCAGCAACATCCCTTGGGCTGTGGCTGAGCACACCAAGAAATTCCTCCCTC 578
QY      76  AGTGGTTGAGCTGTCCATGCCCTGTTCCATTGGAATGCTGAGGAATCTGTCA 135
Db      577  AGTGGTTGAGCTGTCCATGCCCTGTTCCATTGGAATGCTGAGGAATCTGTCA 518
QY      136  ACATGCCACAGCTGCCATGCATTCACATAGCAGCTCATGTTGTCACACATCG 195
Db      517  ACATGCCACAGCTGCCATGCATTCACATAGCAGCTCATGTTGTCACACATCG 458
QY      196  GGTGAGGGGGAGCGCATTCGTCTGAAGCAGCTGGCTGCAAGAGAGCGCTGATTCCA 255

```

```

Db      457  GGTGAGGGGGAGCGCATTCGTCTGAAGGCACTGGCTGCAAGAGCGAGCGTATTCCA 398
QY      256  CCACCGTGGCTGACATGTATCCAAACAAGACTGCAAAATTCAGTGCACACCGAGGCAAG 315
Db      397  CCACCGTGGCTGACATGTATCCAAACAAGACTGCAAAATTCAGTGCACACCGAGGCAAG 338
QY      316  CATGGATCCGCTCGGATGAGATGAGCGGCGAGCGGCTCTCTGTTGCTGCTGCTC 375
Db      337  CATGGATCCGCTCGGATGAGATGAGCGGCGAGCGGCTCTCTGTTGCTGCTGCTC 278
QY      376  CAACACCAAGCATGTGTTTCTGATTGCTCCTCAGTATGGAAAAATTTGTATCTACCAATATC 435
Db      277  CAACACCAAGCATGTGTTTCTGATTGCTCCTCAGTATGGAAAAATTTGTATCTACCAATATC 218
QY      436  CCTGGAGATGGAGAGTGTATGATTCAGTCTATTTATCTTGTGATGCTAATTCCTTGTCT 495
Db      217  CTTGAGAGTGGAGAGTGTATGATTCAGTCTATTTATCTTGTGATGCTAATTCCTTGTCT 158
QY      496  TGTCCCTCAGAAAGTATTCGTTTGTGTTGGGATGAGACAAAGTGGAAATAGAGTGTCTACTA 555
Db      157  TGTCCCTCAGAAAGTATTCGTTTGTGTTGGGATGAGACAAAGTGGAAATAGAGTGTCTACTA 98
QY      556  TATACAGATCATTCGTTGTTAAAGTTTCCAGTTCTGCGAGTTCATGATCTCTGTAATTG 615
Db      97  TATACAGATCATTCGTTGTTAAAGTTTCCAGTTCTGCGAGTTCATGATCTCTGTAATTG 38
QY      616  ATGATGCTGCTGATTCCTACTATTTATCAATCGTCAATTA 652
Db      37  ATGATGCTGCTGATTCAGTCTATTTATCAATCGTCAATTA 1
RESULT 13
LOCUS   AU031690
DEFINITION AU031690 Rice root Oryza sativa (japonica cultivar-group) cDNA
ACCESSION AU031690
VERSION   AU031690.1 GI:3767663
KEYWORDS  EST.
SOURCE    Oryza sativa (japonica cultivar-group)
ORGANISM  Oryza sativa (japonica cultivar-group)
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Eriartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 718)
AUTHORS   Minobe.Y. and Sasaki.T.
TITLE     Rice cDNA from root
JOURNAL   Unpublished
COMMENT   Contact: Takuji Sasaki
          National Institute of Agrobiological Resources
          Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
          305-8602, Japan
          Tel: 81-298-38-7441
          Fax: 81-298-38-7468
          Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
          PROJECT = 'RGP'
          POLVA=No.
FEATURES
    source
        1..718
            /organism="Oryza sativa (japonica cultivar-group)"
            /mol_type="mRNA"
            /cultivar="Nipponbare"
            /db_xref="taxon:39947"
            /clone="R0565_62"
            /clone_lib="Rice root"
            /note="Prepared from seedling root."
BASE COUNT 177 a 158 c 176 g 204 t 3 others
Query Match 89.3%; Score 600.8; DB 9; Length 718;
Best Local Similarity 98.0%; Pred. No. 6.9e-147;
Matches 627; Conservative 0; Mismatches 11; Indels 2; Gaps 2;

```

QY 27 AAATCCATCTTGGGTGGG-GTGAGCACCAAGAAATCTCCCTCAGTGGTTGC 85
Db 52 AAATCCCTCTTGGGTGGGCTTACCAACCAAGAAATCTCCCTCAGTGGTTNC 111
QY 86 AGCTGCCATG-CCGCTGTTCATCTTGGATGCTGAGGAAATCTGTCAACATGCCCA 144
Db 112 ACTGTCCATGCCGCTGTTCATCTTGGATGCTGAGGAAATCTGTCAACATGCCCA 171
QY 145 AGACTCCATGGCAITCACATAGACGCTCCATTTGGTGTGAGCAATCGGTCGAGGG 204
Db 172 AGACTCCATGGCAITCACATAGACGCTCCATTTGGTGTGAGCAATCGGTCGAGGG 231
QY 205 CGAGGCGCATCTCTGGAAGGCACTGGCTGCAAGAGCGACGCTGATTCACACCGCTGG 264
Db 232 CGAGGCGCATCTCTGGAAGGCACTGGCTGCAAGAGCGACGCTGATTCACACCGCTGG 291
QY 265 CTGACATGATCAAAACAGACCTGCAAAATGCAAGTGCACACCGAGGCAAGGATGGATC 324
Db 292 CTGACATGATCAAAACAGACCTGCAAAATGCAAGTGCACACCGAGGCAAGGATGGATC 351
QY 325 CGCTCGGATGAAGATGATGGCGGAGCGGCTTCTGGTGTGCTGCTGCTCAACACCAA 384
Db 352 CGCTCGGATGAAGATGATGGCGGAGCGGCTTCTGGTGTGCTGCTGCTCAACACCAA 411
QY 385 GCATGTGTTCTGATGCTCAGTATGGAATTTGATCTACAGTATCCCTGGAGAG 444
Db 412 GCATGTGTTCTGATGCTCAGTATGGAATTTGATCTACAGTATCCCTGGAGAG 471
QY 445 TGGAGAGTTCATATGAGTCTATTTTATCTGTGATGTAATTCCTTGTCTCCCTCA 504
Db 472 TGGAGAGTTCATATGAGTCTATTTTATCTGTGATGTAATTCCTTGTCTCCCTCA 531
QY 505 GAAGTATCTGTTGTTGGGATGAGCAAGTGGAAATGAGTGTCTATATACACCA 564
Db 532 GAAGTATCTGTTGTTGGGATGAGCAAGTGGAAATGAGTGTCTATATACACCA 591
QY 565 TCATTCGTTGTTAAAGTTTCCAGTCTGAGTTCATGATCTGTAATTTGATGATGCTG 624
Db 592 TCATTCGTTGTTAAAGTTTCCAGTCTGAGTTCATGATCTGTAATTTGATGATGCTG 651
QY 625 GATTTCTATTTATCAATCGCATATATCTGTGTGTA 664
Db 652 GATTTCTATTTATCAATCGCATATATCTGTGTGTA 691

RESULT 14
BQ06858 622 bp mRNA linear EST 19-AUG-2002
LOCUS M020F01 Oryza sativa mature leaf library induced by M.grisea Oryza
DEFINITION sativa cDNA clone M020F01, mRNA sequence.
ACCESSION BQ06858
VERSION BQ06858.1 GI:22305636
KEYWORDS EST.
SOURCE Oryza sativa
ORGANISM Oryza sativa
REFERENCE Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
JOURNAL Ehrhartoideae; Cryzeae; Oryza.
TITLE 1 (bases 1 to 622)
COMMENT D.H.T., Li, D.B., Zhuang, X.F., Dai, C.G., Sun, L.X., Pei, Y.X., Wu
H.F., Jiang, Y.X., Yu, F.C., Gao, Q.K. and Lou, Y.C.
A Gene Expression Screen in Oryza sativa
Unpublished
Contact: Dong HT
Laboratory of Functional Genetics
Bio-technology Institute of Zhejiang University
Kaixun Road 268#, Hangzhou, Zhejiang, P.R.China
Tel: 0086-571-86892051
Fax: 0086-571-86891525
Email: htdong@zjuem.zju.edu.cn
Seq primer: M13 forward primer.
Location/Qualifiers
1. 622
source

/organism="Oryza sativa"
/mol_type="mRNA"
/db_xref="caxon:4530"
/clone="M020F01"
/tissue_type="leaf"
/dev_stage="Mature stage"
/clone_lib="Oryza sativa mature leaf library induced by
M.grisea"
/note="Vector: DSPORT2"
BASE COUNT 154 a 130 c 159 g 179 t
ORIGIN
Query Match 85.2%; Score 573.6; DB 13; Length 622;
Best Local Similarity 98.2%; Pred. No. 9.4e-140;
Matches 601; Conservative 0; Mismatches 9; Indels 2; Gaps 2;
QY 33 CATCTGGGCTGGCTGAGCACACCAAGAAATCTCCCTCAGTGGTTGCAGCTGTC 92
Db 3 CCTCTGGGTTTTGGCTTGAGCACACCAAGAAATCT-CCCTCAGTGGTTGCAGCTGTC 61
QY 93 CAT-GCCGCTTCTCCATTCATTTGGAATGCTGAGAAATCTCTCAATGCCAAGACTGC 151
Db 62 CATGCCGCTTCTCCATTCATTTGGAATGCTGAGAAATCTCTCAATGCCAAGACTGC 121
QY 152 CATGGCATTCACCATAGCAGCTCCATTTGTTGGTGCAGCAATCGGTCGAGGGCGGAGCG 211
Db 122 CATGGCATTCACCATAGCAGCTCCATTTGTTGGTGCAGCAATCGGTCGAGGGCGGAGCG 181
QY 212 CATTCGCTCTGAAGGCACTGGCTGCAAGAGCGAGCTGATTCACACCGCTGGTGACAT 271
Db 182 CATTCGCTCTGAAGGCACTGGCTGCAAGAGCGAGCTGATTCACACCGCTGGTGACAT 241
QY 272 GTATCCAAACAGACTGCAAAATTCAGTGACACCGAGGCAAGGATCCGCTCGC 331
Db 242 GTATCCAAACAGACTGCAAAATTCAGTGACACCGAGGCAAGGATCCGCTCGC 301
QY 332 GATGAAGATGATGGCGGAGCGGCTTCTGGTGGTGGTCTGCTCCACCAAGCATGTG 391
Db 302 GATGAAGATGATGGCGGAGCGGCTTCTGGTGGTGGTCTGCTCCACCAAGCATGTG 361
QY 392 TTTCTGATTTGCTCAGTATGGAATTTGATCTACAGTATCCCTCGAGAGTGGAGAG 451
Db 362 TTTCTGATTTGCTCAGTATGGAATTTGATCTACAGTATCCCTCGAGAGTGGAGAG 421
QY 452 TTGATATTGAGTCTATTTTATCTTGTGATGTAATTTGCTTGTCCCTCAGAGTAT 511
Db 422 TTGATATTGAGTCTATTTTATCTTGTGATGTAATTTGCTTGTCCCTCAGAGTAT 481
QY 512 TCGTTGTTGTTGGATGAGCAAGTGAATAAGAGTCTACTATATACAGATCATTTCT 571
Db 482 TCGTTGTTGTTGGGATGAGCAAGTGAATAAGAGTCTACTATATACAGATCATTTCT 541
QY 572 GTTGTTAAAGTTGCCAGTTCTGCAAGTTCATGATCTGTAATTTGATGTCGAGTTCT 631
Db 542 GTTGTTAAAGTTGCCAGTTCTGCAAGTTCATGATCTGTAATTTGATGTCGAGTTCT 601
QY 632 ACTATTATCAA 643
Db 602 ACTATTAAAAA 613
RESULT 15
CB619892/c
LOCUS OSIIIEa04M03.r OSIIIEa Oryza sativa (indica cultivar-group) cDNA
DEFINITION clone OSIIIEa04M03 3', mRNA sequence.
ACCESSION CB619892
VERSION CB619892.1 GI:29614879
KEYWORDS EST.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzeae; Oryza.
1 (bases 1 to 850)
REFERENCE Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
AUTHORS Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
TITLE Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
JOURNAL Unpublished
COMMENT Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR Primers
FORWARD: gta aaa cga cgg cca gtc
BACKWARD: gga aac agc tat gac cat g
Plate: 04 row: M column: 03
Seq primer: gga aac agc tat gac cat g.

FEATURES

source

FEATURES

source

1..850
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"
/cultivar="IR36"
/db_xref="taxon:39946"
/clone="OSIRI04M03"
/tissue_type="Leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="OSIRI04"
/notes="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
XhoI; Lesion Mimic SPL 11"
BASE COUNT 218 a 232 c 206 g 194 t
ORIGIN

Query Match 84.0%; Score 565.6; DB 14; Length 850;
Best Local Similarity 99.1%; Pred. No. 1.3e-137;
Matches 579; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
QY 17 GGATGGCAGCAAAAC-TCCATCTTTGGGTGTGGGTGAGCACCACCAAGAAATTCCTCCCTC 75
DB 584 GAATGGCAGCAAAACATCCCTCTTGGGTGTGGGTGAGCACCACCAAGAAATTCCTCCCTC 525
QY 76 AGTGGTTTGGAGTGTCCATCGCTGTTCCATTCATGGAAATCTGAGAAATCTGTCA 135
DB 524 AGTGGTTTGGAGTGTCCATCGCTGTTCCATTCATGGAAATCTGAGAAATCTGTCA 465
QY 136 ACATGCCCAAGACTGGCATGGCATTCCACATAGCAGCTCCCATTTGTTGTCAGACAAATCG 195
DB 464 ACATGCCCAAGACTGGCATGGCATTCCACATAGCAGCTCCCATTTGTTGTCAGACAAATCG 405
QY 196 GGTGAGGGGGGAGCGCATTCGTCTGAGGCATCTGCTGCAAGAGCGACGCTGATTCCA 255
DB 404 GGTGAGGGGGGAGCGCATTCGTCTGAGGCATCTGCTGCAAGAGCGACGCTGATTCCA 345
QY 256 CCACCGTGGTGCATGTATCCAAACAAAGACTGCAAAATTCAGTGACACCGAGGGCAAGG 315
DB 344 CCACCGTGGTGCATGTATCCAAACAAAGACTGCAAAATTCAGTGACACCGAGGGCAAGG 285
QY 316 CATGGGATCCGCTCGGATGAAGATGATGCGGGAGCGGGCTTCGTGTGTGTGTGCTC 375
DB 284 CATGGGATCCGCTCGGATGAAGATGATGCGGGAGCGGGCTTCGTGTGTGTGTGCTC 225
QY 376 CAACACCAAGCATGTGTTCTGATTTGCTACTGATTTGGAAATTTCTATCTACCATATC 435
DB 224 CAACACCAAGCATGTGTTCTGATTTGCTACTGATTTGGAAATTTCTATCTACCATATC 165
QY 436 CCTGGAGAGTGGAGAGTGTATTTGATGCTATTTTATCTTGTGATGTAATTCCTTTGCT 495
DB 164 CCTGGAGAGTGGAGAGTGTATTTGATGCTATTTTATCTTGTGATGTAATTCCTTTGCT 105
QY 496 TGTCCTCAGCAAGTATTCGTTTGTGTTGTGGATGAGACAAAGTGGAAATAGAGTGTACTA 555

DB 104 TGTCCTCAGCAAGTATTCGTTTGTGTTGGGTGAGACAAAGTGGAAATAGAGTGTACTA 45
QY 556 TATACACCATCATCTCTGTTGTTAAAGTTTGGCCAGTTCTGCAGTTC 599
DB 44 TATACACCATCATCTCTGTTGTTAAAGTTTGGCCAGTTCTGCAGTTC 1

Search completed: November 14, 2003, 09:20:01
Job time : 1945 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 13:21:35 ; Search time 41 Seconds
(without alignments)
553.607 Million cell updates/sec

Title: US-09-868-025-2

Perfect score: 143

Sequence: 1 MVDNFPISQSTHAWCMS.....LGVLTTPKQWFAAIRAGKV 143

Scoring table: OIIGO

Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A.GeneSeq 19Jun03.*
1: /SID1/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT.*
2: /SID1/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT.*
3: /SID1/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT.*
4: /SID1/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT.*
5: /SID1/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT.*
6: /SID1/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT.*
7: /SID1/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT.*
8: /SID1/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT.*
9: /SID1/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT.*
10: /SID1/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT.*
11: /SID1/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT.*
12: /SID1/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT.*
13: /SID1/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT.*
14: /SID1/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT.*
15: /SID1/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT.*
16: /SID1/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT.*
17: /SID1/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT.*
18: /SID1/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT.*
19: /SID1/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT.*
20: /SID1/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT.*
21: /SID1/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT.*
22: /SID1/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT.*
23: /SID1/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT.*
24: /SID1/gcgdata/geneSeq/geneSeq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	143	100.0	143	22	Rice AGT-SAL 11 pr
2	8	5.6	104	22	Prionibacterium
3	8	5.6	271	22	Drosophila melanog
4	8	5.6	1431	23	Rat Hyaluronic Aci
5	8	5.6	1431	24	Rat Hyaluronic Aci
6	7	4.9	20	22	Albumin-binding pe
7	7	4.9	20	22	Human albumin-bind
8	7	4.9	20	22	Human albumin-bind
9	7	4.9	20	22	Human albumin-bind

10	7	4.9	29	22	AAU14700	Novel bone marrow
11	7	4.9	68	22	AAU65265	Propionibacterium
12	7	4.9	96	23	AAU83203	Novel secreted pro
13	7	4.9	113	22	AAU91935	Human digestive sy
14	7	4.9	113	22	AAU91935	Human colorectal c
15	7	4.9	113	23	ABG97621	Human colorectal c
16	7	4.9	117	23	ABP00108	Human ORFX protein
17	7	4.9	117	22	AAU01584	Human polypeptide
18	7	4.9	283	23	ABZ27335	Streptococcus poly
19	7	4.9	314	22	ABZ71965	Human olfactory re
20	7	4.9	355	22	AAU91032	C glutamicum prote
21	7	4.9	355	23	ABU48300	Listeria monocytog
22	7	4.9	359	22	AAU51254	Human CD72 protein
23	7	4.9	398	20	AAU52838	Escherichia coli p
24	7	4.9	398	23	AAU21878	Isoprenoid related
25	7	4.9	398	23	AAU21885	Isoprenoid related
26	7	4.9	419	23	ABZ28128	Streptococcus poly
27	7	4.9	429	21	AAU14850	Arabidopsis thalia
28	7	4.9	429	21	AAU14850	Arabidopsis thalia
29	7	4.9	450	20	AAU30644	A. thaliana xylan
30	7	4.9	452	21	AAU14049	Arabidopsis thalia
31	7	4.9	452	21	AAU51985	Arabidopsis thalia
32	7	4.9	473	21	AAU51984	Arabidopsis thalia
33	7	4.9	474	21	AAU14048	Arabidopsis thalia
34	7	4.9	481	21	AAU22093	Arabidopsis thalia
35	7	4.9	554	21	AAU22092	Arabidopsis thalia
36	7	4.9	574	21	AAU22091	Arabidopsis thalia
37	7	4.9	633	24	ABP77100	N. gonorrhoeae ami
38	7	4.9	774	19	AAU35390	Plavobacterium odo
39	7	4.9	791	22	ABU23210	Novel human diagno
40	7	4.9	903	23	ABU47324	Listeria monocytog
41	7	4.9	1662	24	AAU32078	Human TRICH-12 pro
42	7	4.9	2516	23	AAU71371	Human ABCA12 prote
43	7	4.9	2595	23	AAU71370	Human ABCA12 prote
44	6	4.2	10	22	AAU68653	Saccharomyces cere
45	6	4.2	12	20	AAU08396	Cysteine mouse lib

ALIGNMENTS

RESULT 1
ID AAB75128 standard; Protein; 143 AA.
XX AAB75128;
XX 03-AUG-2001 (first entry)
DT
DE
DE
XX Rice AGT-SAL 11 protein sequence SEQ ID NO:2.
XX Rice AGT-SAL 11; salt tolerance; plant; cytosolic; antiHIV;
XX Proteinase inhibitor; cancer; human immunodeficiency virus;
XX HIV infection; animal disorder; food processing; enzyme industry;
XX biological preservative.
XX
XX Oryza sativa.
XX
XX WO200130990-A2.
XX
XX 03-MAY-2001.
XX
XX 11-OCT-2000; 2000WO-IN00099.
XX
XX 13-OCT-1999; 99IN-0000997.
XX
XX (AVES-) AVESTAGEN GRAINE TECHNOLOGIES PVT LTD.
XX
XX Patell VM, Antony CM, Chandran D, Madurappa A;
XX WPI; 2001-308632/32.
XX N-PSDB; AAH19867.
XX

PT Nucleotide sequence encoding an AGT-SAL 11 polypeptide similar to
 PT Bowman Birk II type proteinase inhibitors is useful to confer salt
 XX resistance to plants
 XX
 PS Claim 2; Page 17; 22pp; English.

CC The present sequence represents the rice AGT-SAL 11 protein. The
 CC AGT-SAL 11 protein can be used to confer salt tolerance to plants and
 CC other organisms. The AGT-SAL 11 gene was isolated from salt-stressed
 CC rice. Also described are: (i) a transgenic plant comprising a recombinant
 CC expression cassette comprising a plant promoter operably linked to N1;
 CC and (ii) conferring salt tolerance on a plant, comprising introducing
 CC the above expression cassette. The AGT-SAL 11 protein is a proteinase
 CC inhibitor. The proteinase inhibitor may be used to confer stress
 CC tolerance to many plants and organisms including cotton, maize, rice,
 CC soybean, sugar beet, wheat, fruit, vegetables and vines, particularly
 CC biotic bacterial, fungal and pest stresses. Proteinase inhibitors are
 CC also useful in the treatment of cancer, human immunodeficiency virus
 CC (HIV) infection and other animal disorders. The gene may also be useful
 CC in food processing and other enzyme industries as an inhibitor of proteinase
 CC activity and as a biological preservative.

XX Sequence 143 AA;

Query Match 100.0%; Score 143; DB 22; Length 143;
 Best Local Similarity 100.0%; Pred. No. 3.3e-136;
 Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDNTPFIEQSTHAWCWSSTTRPSRHHLRERIPCLALGVTAICSLVWVHSHGGG 60
 DB 1 MVDNTPFIEQSTHAWCWSSTTRPSRHHLRERIPCLALGVTAICSLVWVHSHGGG 60
 QY 61 ISVALCSQCLQTNALRPDPCLTNGGCGYGECHSLGHVDRFPQHSNWNQMDCKPLR 120
 DB 61 ISVALCSQCLQTNALRPDPCLTNGGCGYGECHSLGHVDRFPQHSNWNQMDCKPLR 120
 QY 121 GFPLGLVLTTPKMEFAAIRAGKV 143
 DB 121 GFPLGLVLTTPKMEFAAIRAGKV 143

RESULT 2

AAU48166
 ID AAU48166 standard; Protein; 104 AA.

AC AAU48166;

XX 27-FEB-2002 (first entry)

XX Propionibacterium acnes immunogenic protein #9062.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 XX dermatological; osteopathic; neuroprotectant.

OS Propionibacterium acnes.

XX WO200181581-A2.

XX 01-NOV-2001.

XX 20-APR-2001; 2001WO-US12865.

XX 21-APR-2000; 2000US-199047P.

XX 02-JUN-2000; 2000US-208841P.

XX 07-JUL-2000; 2000US-216747P.

XX (CORI-) CORIXA CORP.

XX Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;

PI L'Maisonneuve J, Zhang Y, Jen S, Carter D;

XX

DR WPI; 2001-616774/71.
 DR N-PSDB; AAS59542.

XX Propionibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris.

XX Example 1; SEQ ID NO 9361; 1069pp; English.

PS Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
 XX polypeptides. The proteins and their associated DNA sequences are used in
 CC the treatment, prevention and diagnosis of medical conditions caused by
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
 CC P. acnes is also involved in infections of bone, joints and the central
 CC nervous system, however it is particularly involved in the inflammatory
 CC lesions associated with acne vulgaris. A method for detecting the
 CC presence or absence of P. acnes in a patient comprises contacting a
 CC sample with a binding agent that binds to the proteins of the invention
 CC and determining the amount of bound protein in the sample. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC specific for P. acnes proteins. These antibodies can be used to
 CC downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA).
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 104 AA;

Query Match 5.6%; Score 8; DB 22; Length 104;
 Best Local Similarity 100.0%; Pred. No. 4.2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 SSTRSPS 28
 DB 47 SSTRSPS 54

RESULT 3

ABB65308

ID ABB65308 standard; Protein; 271 AA.

AC ABB65308;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 22716.

XX Drosophila: developmental biology; cell signalling; insecticide;
 KW pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li FWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL09411.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT Genes from Drosophila and for elucidating cell signalling and cell-cell
 TT interactions -
 PS Disclosure; SEQ ID NO 22716; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (AB57737-AB572072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 271 AA;
 Query Match 5.6%; Score 8; DB 22; Length 271;
 Best Local Similarity 100.0%; Pred. No. 9.5; 0; Indels 0; Gaps 0;
 Matches 8; Conservative 0; Mismatches 0;

QY 21 SSTTRSPS 28
 |||||
 DB 233 SSTTRSPS 240

RESULT 4
 AAM47675
 ID AAM47675 standard; Protein; 1431 AA.

XX AC AAM47675;
 XX DT 22-FEB-2002 (first entry)
 XX DE Rat Hyaluronic Acid Receptor for Endocytosis, HARE.

XX HARE; Hyaluronic Acid Receptor for Endocytosis; hyaluronan; chondroitin;
 KW chondroitin sulphate; extracellular matrix; cartilage; skin;
 KW vitreous humour; endocytic receptor; glycosaminoglycan; rat.
 XX OS Rattus norvegicus.

XX WO2000181544-A2.

XX PD 01-NOV-2001.

XX PF 25-APR-2001; 2001WO-US13403.

XX PR 25-APR-2000; 2000US-199538P.

XX PR 02-NOV-2000; 2000US-245320P.

XX PA (WEIG/) WEIGEL P A.

XX PA (ZHOU/) ZHOU B.

XX PA (WEIG/) WEIGEL J A.

XX PI Weigel PA, Zhou B, Weigel JA;

XX WPI; 2002-049271/06.

XX DR N-PSDB; ABA04648.

XX PT New mammalian hyaluronic acid receptor for endocytosis, useful e.g. for
 PT identifying agents that inhibit binding to hyaluronic acid, and related
 PT nucleic acid -

XX PS Claim 20; Fig 21; 263pp; English.

XX CC The present invention relates to sequences for rat and human HARE
 CC (Hyaluronic Acid Receptor for Endocytosis, ABA04648, ABA04662, AAM47675
 CC and AAM47684). HARE can bind specifically to at least one of hyaluronic
 CC acid (HA, also known as hyaluronan), chondroitin (C) or chondroitin
 CC sulphate (CS). HA is an extracellular matrix component of all tissues,

CC in particular cartilage, skin and vitreous humour. HARE is the endocytic
 CC receptor responsible for removing HA and other glycosaminoglycans from
 CC the circulation. The present sequence is rat HARE.

XX SQ Sequence 1431 AA;

Query Match 5.6%; Score 8; DB 23; Length 1431;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 CLTNNGGC 88
 |||||
 DB 423 CLTNNGGC 430

RESULT 5
 ABG72498
 ID ABG72498 standard; Protein; 1431 AA.

XX AC ABG72498;

XX DT 19-FEB-2003 (first entry)

XX DE Rat 175kDa Hyaluronan receptor for endocytosis (HARE).

XX Hyluronan receptor for endocytosis; HARE; hyaluronan; HA; CD;
 KW chondroitin molecule; chondroitin sulphate; CDS; targeting;
 KW metastasis; tumour; gene therapy; rat; receptor.

XX OS Rattus sp.

XX WO200286093-A2.

XX PD 31-OCT-2002.

XX PF 25-APR-2002; 2002WO-US13209.

XX PR 25-APR-2001; 2001US-286468P.

XX PR 25-APR-2001; 2001US-0842930.

XX PA (WEIG/) WEIGEL P H.

XX PA (WEIG/) WEIGEL J A.

XX PI Weigel PH, Weigel JA;

XX WPI; 2003-0931126/08.

XX DR N-PSDB; ABX13821.

XX PT Targeting compounds e.g. chemotherapeutic agent to cell of subject
 PT expressing functional active hyaluronan receptor for endocytosis of
 PT HARE, by using hyaluronan, HARE, and/or monoclonal antibody that binds
 PT to HARE epitope -

XX Example; Fig 2; 167pp; English.

XX The invention describes a method of targeting a compound to a cell or
 XX tissue of an individual expressing a functionally active hyaluronan (HA)
 XX receptor for endocytosis (HARE) or a cell that does not express
 XX functionally active HARE. The method involves using HA molecule, a
 XX chondroitin molecule (CD), a chondroitin sulphate (CDS) molecule, and/or
 XX a monoclonal antibody raised against a HA-binding domain of HARE.
 XX The method is useful for targeting a compound, preferably a
 XX chemotherapeutic agent or a radioisotope to cell of an individual,
 XX especially a human, expressing HARE on its surface (e.g. gene therapy).
 XX Also described is a method useful for preventing interaction between a
 XX cell having at least one of a HA coat, a CD coat and CDS coat and a cell
 XX expressing HARE on its surface. This second method is useful for
 XX preventing metastasis by preventing interaction between tumour cells
 XX having HA, CD or CDS coat and non-tumour cells expressing HARE on its
 XX surface. The invention also describes a method useful for detecting the
 XX presence of HA, CD and CDS in a biological fluid. This is the amino acid
 XX sequence of the rat 175kDa Hyaluronan receptor for endocytosis (HARE).

XX The present invention relates to novel peptide ligands with affinity for
CC immunoglobulin (Ig) G or serum albumin. The peptide ligands may be
CC conjugated to, and used to prolong the elimination half-time of,
CC therapeutic and diagnostic agents in the body. The present sequence is a
CC human albumin-binding peptide used to illustrate the invention.

XX SQ Sequence 20 AA;
Query Match 4.9%; Score 7; DB 22; Length 20;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 SCXPLRG 121
DB 14 SCXPLRG 20

RESULT 9
ABB44960
ID ABB44960 standard; Peptide; 20 AA.
XX
AC ABB44960;
XX
DT 06-FEB-2002 (first entry)
XX
DE Human albumin-binding peptide #9.
XX
KW Peptide ligand; immunoglobulin G; IgG; serum albumin; human.
XX
OS Synthetic.
XX
PN WO200145746-A2.
XX
PD 28-JUN-2001.
XX
PF 22-DEC-2000; 2000WO-US35325.
XX
PR 23-DEC-1999; 99US-0173048.
XX
PA (GETH) GENENTECH INC.
XX
PI Delano WL, Dennis MS, Lowman HB;
XX
DR WPI; 2001-514266/56.
XX
PT Peptide ligands with affinity for immunoglobulin (Ig) G, IgM and/or
PT human serum albumin which may be conjugated to, and used to prolong the
PT elimination half-time of active agents from the circulation -
XX
PS Example 6; Fig 5; 71pp; English.

XX The present invention relates to novel peptide ligands with affinity for
CC immunoglobulin (Ig) G or serum albumin. The peptide ligands may be
CC conjugated to, and used to prolong the elimination half-time of,
CC therapeutic and diagnostic agents in the body. The present sequence is a
CC human albumin-binding peptide used to illustrate the invention.

XX SQ Sequence 20 AA;
Query Match 4.9%; Score 7; DB 22; Length 20;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 SCXPLRG 121
DB 14 SCXPLRG 20

RESULT 10
AAU14700
ID AAU14700 standard; Protein; 29 AA.
XX

AC AAU14700;
XX
DT 24-OCT-2001 (first entry)
XX
DE Novel bone marrow polypeptide #99.
XX
KW Bone marrow; diagnostic; therapeutic; gene therapy; antigenic;
KW haematopoiesis; myeloid; lymph cell disorder; tissue regeneration;
KW wound healing; nutritional supplement; immune disorder;
XX severe combined immunodeficiency; SCID.
OS Homo sapiens.
XX
PN WO200157187-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US03782.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 30-NOV-2000; 2000US-0250683.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Zhou P, Xue AJ;
PI Ren F, Drmanac RT;
XX
DR WPI; 2001-488875/53.
DR N-PSDB; AAG23005.
XX
PT Nucleic acids encoding bone marrow polypeptides, useful in diagnostic
PT and gene therapy -
XX
PS Claim 10; Page 127; 392pp; English.

XX AAU14602-AAU14794 represent novel bone marrow polypeptides of the
CC invention. The proteins and corresponding coding sequences may be used
CC in the prevention, diagnosis and treatment of diseases associated with
CC inappropriate bone marrow polypeptide expression. For example, to treat
CC disorders associated with decreased expression by rectifying mutations
CC or deletions in a patient's genome that affect the activity of the
CC polypeptides by expressing inactive proteins or to supplement the
CC patient's own production of the polypeptide. Additionally, the nucleic
CC acids may be used to produce the polypeptides, by inserting the nucleic
CC acids into a host cell and culturing the cell to express the protein.
CC The nucleic acid and its complementary sequences may also be used as
CC probes in diagnostic assays to detect and quantitate the presence of
CC similar nucleic acid sequences in samples, and therefore which patients
CC may be in need of restorative therapy. The proteins may also be used as
CC antigens in the production of antibodies against bone marrow proteins.
CC The anti-bone marrow protein antibodies and antagonists may also be used
CC to down regulate expression and activity. The antibodies may also be used
CC as diagnostic agents for detecting the presence of the protein in samples
CC (e.g. by enzyme linked immunosorbent assay (ELISA)). The proteins
CC may be used to regulate haematopoiesis activity, and consequently in the
CC treatment of myeloid or lymph cell disorders; in tissue regeneration,
CC such as wound healing; as a nutritional supplement; and in treatment of
CC immune disorders such as severe combined immunodeficiency (SCID).

XX SQ Sequence 29 AA;
Query Match 4.9%; Score 7; DB 22; Length 29;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 LGVTAIC 48
DB 19 LGVTAIC 25

RESULT 11
AAU65265
ID AAU65265 standard; Protein; 68 AA.
XX AC AAU65265;
XX DT 27-FEB-2002 (first entry)
XX DE Propionibacterium acnes immunogenic protein #26161.
XX KW SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;
XX KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
XX KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
XX KW dermatological; osteopathic; neuroprotectant.
XX OS Propionibacterium acnes.
XX PN WO2000181581-A2.
XX PD 01-NOV-2001.
XX PF 20-APR-2001; 2001WO-US12865.
XX PR 21-APR-2000; 2000US-199047P.
XX PR 02-JUN-2000; 2000US-208841P.
XX PR 07-JUL-2000; 2000US-216747P.
XX PA (CORI-) CORIXA CORP.
XX PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
XX PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX DR WPI; 2001-616774/71.
XX DR N-PSDB; AAGS9663.
XX PT Propionibacterium acnes polypeptides and nucleic acids useful for
XX PT vaccinating against and diagnosing infections, especially useful for
XX PT treating acne vulgaris.
XX Example 1; SEQ ID No 26460; 1069pp; English.
XX CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
XX CC polypeptides. The proteins and their associated DNA sequences are used in
XX CC the treatment, prevention and diagnosis of medical conditions caused by
XX CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
XX CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
XX CC P. acnes is also involved in infections of bone, joints and the central
XX CC nervous system, however it is particularly involved in the inflammatory
XX CC lesions associated with acne vulgaris. A method for detecting the
XX CC presence or absence of P. acnes in a patient comprises contacting a
XX CC sample with a binding agent that binds to the proteins of the invention
XX CC and determining the amount of bound protein in the sample. The
XX CC polypeptides may be used as antigens in the production of antibodies
XX CC specific for P. acnes proteins. These antibodies can be used to
XX CC downregulate expression and activity of P. acnes polypeptides and
XX CC therefore treat P. acnes infections. The antibodies may also be used as
XX CC diagnostic agents for determining P. acnes presence, for example, by
XX CC enzyme linked immunosorbent assay (ELISA).
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 68 AA;
Query Match 4.9%; Score 7; DB 22; Length 68;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 12
AAU83203
ID AAU83203 standard; Protein; 96 AA.
XX AC AAU83203;
XX DT 08-MAY-2002 (first entry)
XX DE Novel secreted protein Z892026G7P.
XX KW Protein secretion; mammalian secreted polypeptide; MSP.
XX KW Homo sapiens.
XX OS WO2000202621-A2.
XX PN 10-JAN-2002.
XX PF 28-JUN-2001; 2001WO-US20638.
XX PR 30-JUN-2000; 2000US-215446P.
XX PA (ZYMO) ZYMOGENETICS INC.
XX PI Sheppard PO, Presnell SR;
XX DR WPI; 2002-147999/19.
XX DR N-PSDB; ABK33118.
XX PT Novel isolated mammalian secreted polypeptide useful in therapeutic and
XX PT diagnostic methods, to direct secretion of other proteins of interest
XX PT from host cell, as educational tools, and as laboratory practicum kits
XX PS Claim 12; Page 322; 397pp; English.
XX CC The invention describes an isolated mammalian secreted polypeptide (MSP)
XX CC (I). (I) is useful to direct the secretion of other proteins of interest
XX CC from a host cell, to monitor secretion of proteins, to degenerate
XX CC sequences comprising all nucleotide sequences encoding a particular
XX CC polypeptide, to screen for cell metabolism effecting receptors, for
XX CC identifying new target receptors and drug design, for identifying, for
XX CC protein purification, for determining the weight of expressed MSP
XX CC polypeptides as a ratio to total protein expressed, for identifying
XX CC peptide cleavage sites, for coupling amino and carboxy terminal tags, for
XX CC amino acid sequence analysis, for monitoring biological activities of the
XX CC protein in vitro and in vivo, and to teach analytical skills and as
XX CC reagents for the study of cells, receptors, and other binding molecules.
XX CC The polynucleotide is useful for radiation hybrid mapping, and somatic
XX CC cell genetic technique developed for constructing high-resolution,
XX CC contiguous maps of mammalian chromosomes. Reagents disclosed in the
XX CC invention may be used to detect metabolic abnormalities characterised by
XX CC over or under production of the protein. This is the amino acid sequence
XX CC of a mammalian secreted polypeptide, described in the method of the
XX CC invention.
XX SQ Sequence 96 AA;
Query Match 4.9%; Score 7; DB 23; Length 96;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 LRGEFLG 125
DB 33 LRGEFLG 39

QY 61 ISVALCS 67
DB 54 ISVALCS 60
Query Match 4.9%; Score 7; DB 23; Length 96;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
RESULT 13
AAU91935
ID AAU91935 standard; Protein; 113 AA.
XX AC AAU91935;
XX XX

DT 05-NOV-2001 (first entry)
XX Human digestive system antigen SEQ ID NO: 1284.
DE
XX
XX Human, digestive system antigen; gene therapy; cancer; appendicitis;
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KM digestive system disorder; Meckel's diverticulum.
XX
OS Homo sapiens.
XX
XX WO200155314-A2.
XX
PD 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01324.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0236273.
PR 22-AUG-2000; 2000US-0236681.
PR 22-AUG-2000; 2000US-0236868.
PR 22-AUG-2000; 2000US-0237182.
PR 23-AUG-2000; 2000US-0237009.
PR 01-SEP-2000; 2000US-0238924.
PR 01-SEP-2000; 2000US-0239287.
PR 01-SEP-2000; 2000US-0239343.
PR 01-SEP-2000; 2000US-0239344.
PR 01-SEP-2000; 2000US-0239345.
PR 05-SEP-2000; 2000US-0239509.
PR 05-SEP-2000; 2000US-0239513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.

PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249219.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.

PR	11-JUL-2000;	2000US-0217436;
PR	14-JUL-2000;	2000US-0218290;
PR	26-JUL-2000;	2000US-0220963;
PR	26-JUL-2000;	2000US-0220964;
PR	14-AUG-2000;	2000US-0224518;
PR	14-AUG-2000;	2000US-0224519;
PR	14-AUG-2000;	2000US-02245213;
PR	14-AUG-2000;	2000US-0225213;
PR	14-AUG-2000;	2000US-0225214;
PR	14-AUG-2000;	2000US-0225216;
PR	14-AUG-2000;	2000US-0225267;
PR	14-AUG-2000;	2000US-0225268;
PR	14-AUG-2000;	2000US-0225270;
PR	14-AUG-2000;	2000US-0225447;
PR	14-AUG-2000;	2000US-0225757;
PR	14-AUG-2000;	2000US-0225758;
PR	14-AUG-2000;	2000US-0225759;
PR	18-AUG-2000;	2000US-0226279;
PR	18-AUG-2000;	2000US-0226681;
PR	22-AUG-2000;	2000US-0226868;
PR	22-AUG-2000;	2000US-0226869;
PR	23-AUG-2000;	2000US-0227082;
PR	23-AUG-2000;	2000US-0227009;
PR	30-SEP-2000;	2000US-0228924;
PR	01-SEP-2000;	2000US-0229287;
PR	01-SEP-2000;	2000US-0229343;
PR	01-SEP-2000;	2000US-0229344;
PR	01-SEP-2000;	2000US-0229345;
PR	05-SEP-2000;	2000US-0229509;
PR	05-SEP-2000;	2000US-0229513;
PR	06-SEP-2000;	2000US-0230437;
PR	06-SEP-2000;	2000US-0230438;
PR	14-SEP-2000;	2000US-0230398;
PR	14-SEP-2000;	2000US-0230399;
PR	14-SEP-2000;	2000US-0230400;
PR	14-SEP-2000;	2000US-0230401;
PR	14-SEP-2000;	2000US-0230631;
PR	14-SEP-2000;	2000US-0230632;
PR	14-SEP-2000;	2000US-0230635;
PR	21-SEP-2000;	2000US-0234223;
PR	21-SEP-2000;	2000US-0234274;
PR	25-SEP-2000;	2000US-0234997;
PR	25-SEP-2000;	2000US-0234998;
PR	26-SEP-2000;	2000US-0235484;
PR	27-SEP-2000;	2000US-0235834;
PR	27-SEP-2000;	2000US-0235836;
PR	29-SEP-2000;	2000US-0236327;
PR	29-SEP-2000;	2000US-0236367;
PR	29-SEP-2000;	2000US-0236368;
PR	29-SEP-2000;	2000US-0236369;
PR	29-SEP-2000;	2000US-0236370;
PR	02-OCT-2000;	2000US-0236802;
PR	02-OCT-2000;	2000US-0237037;
PR	02-OCT-2000;	2000US-0237038;
PR	02-OCT-2000;	2000US-0237039;
PR	02-OCT-2000;	2000US-0237040;
PR	13-OCT-2000;	2000US-0239935;
PR	13-OCT-2000;	2000US-0240960;
PR	20-OCT-2000;	2000US-0241121;
PR	20-OCT-2000;	2000US-0241785;
PR	20-OCT-2000;	2000US-0241786;
PR	20-OCT-2000;	2000US-0241808;
PR	20-OCT-2000;	2000US-0241809;
PR	20-OCT-2000;	2000US-0241826;

01-NOV-2000; 2000US-0244617.
 08-NOV-2000; 2000US-0246474.
 08-NOV-2000; 2000US-0246475.
 08-NOV-2000; 2000US-0246476.
 08-NOV-2000; 2000US-0246477.
 08-NOV-2000; 2000US-0246478.
 08-NOV-2000; 2000US-0246523.
 08-NOV-2000; 2000US-0246524.
 08-NOV-2000; 2000US-0246525.
 08-NOV-2000; 2000US-0246526.
 08-NOV-2000; 2000US-0246527.
 08-NOV-2000; 2000US-0246528.
 08-NOV-2000; 2000US-0246532.
 08-NOV-2000; 2000US-0246609.
 08-NOV-2000; 2000US-0246609.
 08-NOV-2000; 2000US-0246610.
 08-NOV-2000; 2000US-0246611.
 08-NOV-2000; 2000US-0246613.
 17-NOV-2000; 2000US-0249207.
 17-NOV-2000; 2000US-0249208.
 17-NOV-2000; 2000US-0249209.
 17-NOV-2000; 2000US-0249210.
 17-NOV-2000; 2000US-0249211.
 17-NOV-2000; 2000US-0249212.
 17-NOV-2000; 2000US-0249213.
 17-NOV-2000; 2000US-0249214.
 17-NOV-2000; 2000US-0249215.
 17-NOV-2000; 2000US-0249216.
 17-NOV-2000; 2000US-0249217.
 17-NOV-2000; 2000US-0249218.
 17-NOV-2000; 2000US-0249244.
 17-NOV-2000; 2000US-0249245.
 17-NOV-2000; 2000US-0249265.
 17-NOV-2000; 2000US-0249267.
 17-NOV-2000; 2000US-0249297.
 17-NOV-2000; 2000US-0249299.
 17-NOV-2000; 2000US-0249300.
 01-DEC-2000; 2000US-0250160.
 01-DEC-2000; 2000US-0250391.
 05-DEC-2000; 2000US-0251030.
 05-DEC-2000; 2000US-0251988.
 05-DEC-2000; 2000US-0256719.
 06-DEC-2000; 2000US-0251479.
 08-DEC-2000; 2000US-0251856.
 08-DEC-2000; 2000US-0251868.
 08-DEC-2000; 2000US-0251869.
 08-DEC-2000; 2000US-0251889.
 11-DEC-2000; 2000US-0251990.
 05-JAN-2001; 2001US-0254097.
 05-JAN-2001; 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Barash SC, Ruben SM;
 WPI; 2001-457727/49.
 N-PSDB; AA157547.
 Isolated polypeptide for treating, preventing and/or prognosing
 disorders related to the colon and rectum including colorectal cancers
 and also for testing and detection e.g. diagnosis -
 Claim 11; SEQ ID NO: 84; 522pp + Sequence Listing; English.
 The present invention provides the protein and coding sequences of a
 number of colorectal cancer antigens. These are shown in
 AA157547-AA157619 and AA157619-AA157641. These can be used in the
 diagnosis, prevention and treatment of cancer of the colon and/or rectum.
 The present sequence is a colorectal cancer antigen of the invention.
 Note: The sequence data for this patent did not form part of the printed
 specification, but was obtained in electronic format directly from WIPO
 at ftp.wipo.int/pub/published_pct_sequences.
 Sequence 113 AA;
 SQ

Query Match 4.9%; Score 7; DB 23; Length 113;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 AICSLVW 52
 Db 68 AICSLVW 74
 |||||
 |||||

RESULT 15
 ABG97621
 ID ABG97621 standard; Protein; 113 AA.
 XX
 AC ABG97621;
 XX
 DT 18-DEC-2002 (first entry)
 XX
 DE Human colorectal cancer related protein #1.
 XX
 KW Human; colorectal cancer related protein; colon; rectum;
 KW colorectal cancer metastasis; gastrointestinal disorder;
 KW cytostatic.
 XX
 OS Homo sapiens.
 XX
 PN US2002119919-A1.
 XX
 PD 29-AUG-2002.
 XX
 PF 17-JAN-2001; 2001US-0764855.
 XX
 PR 31-JAN-2000; 2000US-179065P.
 XX
 PA (ROSE/) ROSEN C A.
 PA (RUBE/) RUBEN S M.
 PA (BARA/) BARASH S C.
 XX
 PI Rosen CA, Ruben SM, Barash SC;
 XX
 WPI; 2002-731367/79.
 N-PSDB; ABS99724.
 XX
 PT New colorectal cancer polypeptide for diagnosing, prognosing,
 PT preventing, and treating immune, hyperproliferative, liver, kidney,
 PT reproductive disorders and for identifying modulators of therapeutic
 use -
 XX
 PS Claim 11; SEQ ID No 84; 183pp; English.
 XX
 CC The present invention relates to the isolation of novel human
 CC colorectal cancer related proteins, and polynucleotide sequences
 CC encoding them. The sequences of the invention are useful in the
 CC diagnosis, treatment, prevention and/or prognosis of the colon
 CC and/or rectum, including colorectal cancer, colorectal cancer
 CC metastases, and gastrointestinal disorders such as dysphagia,
 CC peptic oesophagitis, gastric reflux, irritable bowel syndrome,
 CC and peritoneal diseases. The invention also describes antibodies
 CC that bind colorectal cancer related proteins, vectors, host cells,
 CC and recombinant and synthetic methods for producing human colorectal
 CC cancer related polynucleotides, polypeptides, and/or antibodies.
 CC ABG97621-ABG97693 represent human colorectal cancer related proteins.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC USPTO web site at seqdata.uspto.gov/patseqDIDEntry.html.
 XX
 SQ Sequence 113 AA;
 SQ

Query Match 4.9%; Score 7; DB 23; Length 113;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 AICSLVW 52

Db |||||||
 68 AICSLW 74

Search completed: November 13, 2003, 13:27:22
Job time : 43 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 13:26:31 ; Search time 21 Seconds

(without alignments)

288.117 Million cell updates/sec

Title: US-09-868-025-2

Perfect score: 143

Sequence: 1 MVDNPPISQSETHAWCS.....LGVLPKPKMFAAIRAGKV 143

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- Issued Patents AA.*
- 1: /cgn2_6/ptodata/1/iaa/5A COMB.pcp.*
 - 2: /cgn2_6/ptodata/1/iaa/5B COMB.pcp.*
 - 3: /cgn2_6/ptodata/1/iaa/6A COMB.pcp.*
 - 4: /cgn2_6/ptodata/1/iaa/6B COMB.pcp.*
 - 5: /cgn2_6/ptodata/1/iaa/PTCUS COMB.pcp.*
 - 6: /cgn2_6/ptodata/1/iaa/backfiles1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	4.9	102	4	US-09-634-238-332
2	7	4.9	133	4	US-09-252-991A-28337
3	7	4.9	140	4	US-09-252-991A-32132
4	7	4.9	141	4	US-09-252-991A-16711
5	7	4.9	201	4	US-09-252-991A-24176
6	7	4.9	236	4	US-09-252-991A-27371
7	7	4.9	388	4	US-09-252-991A-31365
8	7	4.9	478	4	US-09-252-991A-32557
9	7	4.9	491	4	US-09-252-991A-29642
10	7	4.9	774	3	US-09-346-237-8
11	6	4.2	44	1	US-08-530-010-33
12	6	4.2	44	2	US-08-494-101B-33
13	6	4.2	44	3	US-08-714-524D-33
14	6	4.2	62	4	US-09-134-001C-4040
15	6	4.2	68	4	US-09-107-532A-4813
16	6	4.2	84	3	US-09-386-493-14
17	6	4.2	115	4	US-09-308-246C-2
18	6	4.2	121	3	US-09-171-982-2
19	6	4.2	121	4	US-09-171-982-2
20	6	4.2	123	4	US-09-252-991A-32817
21	6	4.2	130	1	US-08-580-680-1
22	6	4.2	130	1	US-08-480-156A-1
23	6	4.2	130	1	US-08-354-961-1
24	6	4.2	130	2	US-08-581-094-1
25	6	4.2	130	2	US-08-580-665-1
26	6	4.2	130	2	US-08-581-142-1
27	6	4.2	130	2	US-08-944-449-2

Sequence 4, Appli
Sequence 1, Appli
Sequence 2, Appli
Sequence 4, Appli
Sequence 1, Appli
Sequence 28438, A
Sequence 18080, A
Sequence 26530, A
Sequence 18992, A
Sequence 19973, A
Sequence 25898, A
Sequence 27639, A
Sequence 27562, A
Sequence 28681, A
Sequence 17497, A
Sequence 6, Appli
Sequence 27711, A
Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-634-238-332

; Sequence 332, Application US/09634238

; Patent No. 6544772

; GENERAL INFORMATION:

; APPLICANT: Glenn, Matthew

; APPLICANT: Havukkala, Ilkka J.

; APPLICANT: Bloksberg, Leonard, N.

; APPLICANT: Lubbers, Mark W.

; APPLICANT: Dekker, James

; APPLICANT: Christenson, Anna C.

; APPLICANT: Holland, Ross

; APPLICANT: O'Toole, Paul W.

; APPLICANT: Reid, Julian R.

; APPLICANT: Coolbear, Timothy

; TITLE OF INVENTION: Polynucleotides, materials incorporating

; FILE REFERENCE: 11000.1043U1

; CURRENT APPLICATION NUMBER: US/09/634,238

; NUMBER OF SEQ ID NOS: 422

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 332

; LENGTH: 102

; TYPE: PRT

; ORGANISM: Lactobacillus rhamnosus

US-09-634-238-332

Query Match 4.9%; Score 7; DB 4; Length 102;

Best Local Similarity: 100.0%; Pred. No. 7.8;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 ALGVTAI 47

|||||

Db 46 ALGVTAI 52

RESULT 2

US-09-252-991A-28337

; Sequence 28337, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

```
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28337
; LENGTH: 133
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28337

Query Match      4.9%; Score 7; DB 4; Length 133;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      19 WSSSTTR 25
      |||||
Db      105 WSSSTTR 111

RESULT 3
US-09-252-991A-32132
; Sequence 32132, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32132
; LENGTH: 140
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32132

Query Match      4.9%; Score 7; DB 4; Length 140;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      21 SSTTRSP 27
      |||||
Db      24 SSTTRSP 30

RESULT 4
US-09-252-991A-16711
; Sequence 16711, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16711
; LENGTH: 141
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16711

Query Match      4.9%; Score 7; DB 4; Length 141;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      21 SSTTRSP 27
      |||||
Db      24 SSTTRSP 30

RESULT 5
US-09-252-991A-24176
; Sequence 24176, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24176
; LENGTH: 201
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24176

Query Match      4.9%; Score 7; DB 4; Length 201;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      74 ALRPRPD 80
      |||||
Db      146 ALRPRPD 152

RESULT 6
US-09-252-991A-27371
; Sequence 27371, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27371
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27371

Query Match      4.9%; Score 7; DB 4; Length 236;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      20 SSSSTRS 26
      |||||
Db      34 SSSSTRS 40

RESULT 7
US-09-252-991A-31265
; Sequence 31265, Application US/09252991A
; Patent No. 6551795
```

```

; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31265
; LENGTH: 388
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31265

Query Match 4.9%; Score 7; DB 4; Length 388;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 SSTTRSP 27
DB 318 SSTTRSP 324

RESULT 8
US-09-252-991A-32557
; Sequence 32557, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32557
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32557

Query Match 4.9%; Score 7; DB 4; Length 478;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 SRHLHR 34
DB 137 SRHLHR 143

RESULT 9
US-09-252-991A-29642
; Sequence 29642, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32557
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32557

Query Match 4.9%; Score 7; DB 4; Length 478;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 SRHLHR 34
DB 137 SRHLHR 143

RESULT 10
US-09-346-237-8
; Sequence 8, Application US/09346237A
; Patent No. 6265197
; GENERAL INFORMATION:
; APPLICANT: Bisgard-Frantzen, Henrik
; APPLICANT: Svendsen, Allan
; TITLE OF INVENTION: Starch Debranching Enzymes
; FILE REFERENCE: 5629.200-US
; CURRENT APPLICATION NUMBER: US/09/346,237A
; CURRENT FILING DATE: 1999-07-01
; EARLIER APPLICATION NUMBER: PA 1998 00868
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: 60/094,353
; EARLIER FILING DATE: 1998-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 774
; TYPE: PRT
; ORGANISM: Favobacterium odoratum
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)...(774)
; OTHER INFORMATION: Isomylase
US-09-346-237-8

Query Match 4.9%; Score 7; DB 3; Length 774;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 ALGVTAI 47
DB 248 ALGVTAI 254

RESULT 11
US-08-530-010-33
; Sequence 33, Application US/08530010
; Patent No. 5689055
; GENERAL INFORMATION:
; APPLICANT: Meyerowitz, Elliott M.
; APPLICANT: Chang, Caren
; APPLICANT: Bleeker, Anthony B.
; TITLE OF INVENTION: PLANTS HAVING MODIFIED RESPONSE TO ETHYLENE
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Richard F. Trecartin
; STREET: 3400 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 29642
; LENGTH: 491
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29642

Query Match 4.9%; Score 7; DB 4; Length 491;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 136 AAIRAGK 142
DB 255 AAIRAGK 261

RESULT 10
US-09-346-237-8
; Sequence 8, Application US/09346237A
; Patent No. 6265197
; GENERAL INFORMATION:
; APPLICANT: Bisgard-Frantzen, Henrik
; APPLICANT: Svendsen, Allan
; TITLE OF INVENTION: Starch Debranching Enzymes
; FILE REFERENCE: 5629.200-US
; CURRENT APPLICATION NUMBER: US/09/346,237A
; CURRENT FILING DATE: 1999-07-01
; EARLIER APPLICATION NUMBER: PA 1998 00868
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: 60/094,353
; EARLIER FILING DATE: 1998-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 774
; TYPE: PRT
; ORGANISM: Favobacterium odoratum
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)...(774)
; OTHER INFORMATION: Isomylase
US-09-346-237-8

Query Match 4.9%; Score 7; DB 3; Length 774;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 ALGVTAI 47
DB 248 ALGVTAI 254

RESULT 11
US-08-530-010-33
; Sequence 33, Application US/08530010
; Patent No. 5689055
; GENERAL INFORMATION:
; APPLICANT: Meyerowitz, Elliott M.
; APPLICANT: Chang, Caren
; APPLICANT: Bleeker, Anthony B.
; TITLE OF INVENTION: PLANTS HAVING MODIFIED RESPONSE TO ETHYLENE
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Richard F. Trecartin
; STREET: 3400 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/530,010
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA: US 08/086,555
APPLICATION NUMBER: US 08/086,555
FILING DATE: 01-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Trecartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-57515/RFT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 44 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-530-010-33

Query Match 4.2%; Score 6; DB 1; Length 44;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 SGMDS 116
DB 21 SGMDS 26

RESULT 12
US-08-484-101B-33
Sequence 33, Application US/08484101B
Patent No. 5824868
GENERAL INFORMATION:
APPLICANT: California Institute of Technology
TITLE OF INVENTION: PLANTS HAVING MODIFIED RESPONSE TO
TITLE OF INVENTION: ETHYLENE
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: Richard F. Trecartin
STREET: 3400 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,101B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA: PCT/US94/
APPLICATION NUMBER: 01-JUL-1994
FILING DATE: 01-JUL-1994
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/086,555
FILING DATE: 01-JUL-1993
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Trecartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-57515-2/RFT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249

INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 44 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-484-101B-33

Query Match 4.2%; Score 6; DB 2; Length 44;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 SGMDS 116
DB 21 SGMDS 26

RESULT 13
US-08-714-524D-33
Sequence 33, Application US/08714524D
Patent No. 6294716
GENERAL INFORMATION:
APPLICANT: Meyerowitz, Elliott M
APPLICANT: Chang, Caren
APPLICANT: Blescker, Anthony B
TITLE OF INVENTION: PLANTS HAVING MODIFIED RESPONSE TO ETHYLENE
FILE REFERENCE: A-57515-4
CURRENT APPLICATION NUMBER: US/08/714,524D
CURRENT FILING DATE: 1996-09-16
NUMBER OF SEQ ID NOS: 56
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 33
LENGTH: 44
TYPE: PRT
ORGANISM: Escherichia coli
US-08-714-524D-33

Query Match 4.2%; Score 6; DB 3; Length 44;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 SGMDS 116
DB 21 SGMDS 26

RESULT 14
US-09-134-001C-4040
Sequence 4040, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 4040
LENGTH: 62
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4040

Query Match 4.2%; Score 6; DB 4; Length 62;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 TTRSPS 28
 |||||
 Db 6 TTRSPS 11

RESULT 15

US-09-107-532A-4813
 ; Sequence 4813, Application US/09107532A
 ; Patent No. 6583275
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn A Doucette-Stamm and David Bush
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
 ; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street

CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD/ROM ISO9660

COMPUTER: PC

OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Arinello, Pamela Deneke

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 4813:

SEQUENCE CHARACTERISTICS:

LENGTH: 68 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Enterococcus faecium

FEATURE:

NAME/KEY: misc feature

LOCATION: (1B) LOCATION 1...68

SEQUENCE DESCRIPTION: SEQ ID NO: 4813:

US-09-107-532A-4813

Query Match 4.2%; Score 6; DB 4; Length 68;
 Best Local Similarity 100.0%; Pred.No. 59;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 SSTRS 26
 |||||
 Db 48 SSTRS 53

Search completed: November 13, 2003, 13:29:46
 Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2003, 08:04:27 ; Search time 58 Seconds
(without alignments)
5121.566 Million cell updates/sec

Title: US-09-868-025-1

Perfect score: 673
Sequence: 1 ttatctgtctgtctgtctgtat.....actgtgtgtaaaaaaaaaa 673

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38.8	5.8	528	4 US-09-107-532A-991	Sequence 991, App
2	38.8	5.8	552	4 US-09-107-532A-3543	Sequence 3543, App
3	38.8	5.8	1476	4 US-09-107-532A-2461	Sequence 2461, App
4	35.8	5.3	11091	4 US-09-134-001C-2243	Sequence 2243, App
5	35	5.2	882	4 US-09-134-001C-1530	Sequence 1530, App
6	34.6	5.1	4403765	3 US-09-103-840A-2	Sequence 2, Appli
7	34.6	5.1	4411529	3 US-09-103-840A-1	Sequence 1, Appli
8	34	5.1	649	4 US-09-071-035-231	Sequence 231, App
9	34	5.1	834	4 US-09-071-035-229	Sequence 229, App
10	33.6	5.0	3207	1 US-08-045-806-1	Sequence 1, Appli
11	33.6	5.0	3207	1 US-08-366-051B-1	Sequence 1, Appli
12	32.8	4.9	1830121	4 US-09-557-884-1	Sequence 1, Appli
13	32.8	4.9	1830121	4 US-09-643-990A-1	Sequence 1, Appli
14	32.6	4.8	2145	4 US-09-059-584-48	Sequence 48, Appl
15	32.6	4.8	2287	4 US-09-059-584-47	Sequence 47, Appl
16	31.8	4.7	1664976	4 US-08-916-421B-1	Sequence 1, Appli
17	31.6	4.7	518	4 US-09-228-986-17	Sequence 17, Appl
18	31.4	4.7	16382	3 US-08-718-388-8	Sequence 8, Appli
19	31	4.6	4078	4 US-09-016-434-1134	Sequence 1134, Ap
20	31	4.6	4651	2 US-08-443-644-4	Sequence 4, Appli
21	31	4.6	4651	2 US-08-087-244A-4	Sequence 4, Appli
22	30.8	4.6	739	4 US-08-936-165A-166	Sequence 166, App
23	30.8	4.6	2118	2 US-08-619-362A-7	Sequence 7, Appli
24	30.8	4.6	2119	4 US-09-151-771B-7	Sequence 7, Appli
25	30.6	4.5	51259	3 US-08-781-891-209	Sequence 209, App
26	30.6	4.5	51259	4 US-09-618-166-209	Sequence 209, App
27	30.4	4.5	32155	4 US-08-311-731A-1	Sequence 1, Appli

C 28 30.2 4.5 957 4 US-09-252-991A-14974 Sequence 14974, A
C 29 30.2 4.5 1107 4 US-09-252-991A-456 Sequence 456, App
C 30 30.2 4.5 1326 4 US-09-252-991A-409 Sequence 409, App
C 31 30.2 4.5 1579 4 US-09-620-312D-927 Sequence 927, App
C 32 30.2 4.5 1641 4 US-09-252-991A-15102 Sequence 15102, A
C 33 30.2 4.5 3054 4 US-09-484-970B-138 Sequence 138, App
C 34 30.2 4.5 168575 4 US-09-426-290-1 Sequence 1, Appli
C 35 30.2 4.5 1230025 4 US-09-198-452A-1 Sequence 1, Appli
C 36 30 4.5 576 4 US-08-671-548C-1 Sequence 11, Appli
C 37 30 4.5 2749 4 US-09-619-353-11 Sequence 7, Appli
C 38 30 4.5 3661 3 US-08-822-324-7 Sequence 7, Appli
C 39 30 4.5 3561 3 US-08-718-388-5 Sequence 5, Appli
C 40 30 4.5 5340 4 US-09-627-122-21 Sequence 21, Appli
C 41 30 4.5 7824 3 US-08-718-388-6 Sequence 6, Appli
C 42 30 4.5 1830121 4 US-09-557-884-1 Sequence 1, Appli
C 43 30 4.5 1830121 4 US-09-643-990A-1 Sequence 1, Appli
C 44 29.8 4.4 176 4 US-09-702-705-1720 Sequence 1720, Ap
C 45 29.8 4.4 176 4 US-09-736-457-1720 Sequence 1720, Ap

ALIGNMENTS

RESULT 1
US-09-107-532A-991
: Sequence 991, Application US/09107532A
: Patent No. 6583275
: GENERAL INFORMATION:
: APPLICANT: Lynn A Doucette-Stamm and David Bush
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
: NUMBER OF SEQUENCES: 7310
: CORRESPONDENCE ADDRESS:
: ADDRESSER: GENOME THERAPEUTICS CORPORATION
: STREET: 100 Beaver Street
: CITY: Waltham
: STATE: Massachusetts
: COUNTRY: USA
: ZIP: 02354
: COMPUTER READABLE FORM:
: MEDIUM TYPE: CD/ROM ISO9660
: COMPUTER: PC
: OPERATING SYSTEM: <Unknown>
: SOFTWARE: ASCII
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/107,532A
: FILING DATE: 30-Jun-1998
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/085,598
: FILING DATE: 14 May 1998
: APPLICATION NUMBER: 60/051571
: FILING DATE: July 2, 1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Ariniello, Pamela Deneka
: REGISTRATION NUMBER: 40,489
: REFERENCE/DOCKET NUMBER: GTC-012
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (781)893-5007
: TELEFAX: (781)893-8277
: INFORMATION FOR SEQ ID NO: 991:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 528 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: circular
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Enterococcus faecium
: FEATURE:
: NAME/KEY: misc.feature
: LOCATION: (B) LOCATION 1...528

```
SEQUENCE DESCRIPTION: SEQ ID NO: 991;
US-09-107-532A-991
Query Match 5.8%; Score 38.8; DB 4; Length 528;
Best Local Similarity 60.4%; Pred. No. 0.018; Indels 0; Gaps 0;
Matches 64; Conservative 0; Mismatches 42;

QY 430 AGTATCCCTGGAGAGTGGAGAGTTGATATTCAGTCTATTTTATCTTGTGATGTAATGGC 489
DB 19 AGTATTGCTTCTGATTCGAAAGATTCTGTTGATCTATTTGCTGTGGTGCAGAAAAGTT 78
QY 490 TTGCTTGTCCCTCAGAAAGTATTCGTTTGTGTTGGGATGAGACAA 535
DB 79 TTGTTGTCTCCCTCCGTTTTTGGCGCTTCCTTTTGAAGATTCTA 124

RESULT 3
US-09-107-532A-2461/C
; Sequence 2461; Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 2461:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1476 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...1476
; SEQUENCE DESCRIPTION: SEQ ID NO: 2461:
US-09-107-532A-2461
Query Match 5.8%; Score 38.8; DB 4; Length 1476;
Best Local Similarity 60.4%; Pred. No. 0.031;
Matches 64; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 430 AGTATCCCTGGAGAGTGGAGAGTTGATATTCAGTCTATTTTATCTTGTGATGTAATGGC 489
DB 554 AGTATTGCTTCTGATTCGAAAGATTCTGTTGATCTATTTGCTGTGGTGCAGAAAAGTT 495
QY 490 TTGCTTGTCCCTCAGAAAGTATTCGTTTGTGTTGGGATGAGACAA 535
DB 494 TTGTTGTCTCCCTCCGTTTTTGGCGCTTCCTTTTGAAGATTCTA 419
```

```
SEQUENCE DESCRIPTION: SEQ ID NO: 991;
US-09-107-532A-991
Query Match 5.8%; Score 38.8; DB 4; Length 528;
Best Local Similarity 60.4%; Pred. No. 0.018; Indels 0; Gaps 0;
Matches 64; Conservative 0; Mismatches 42;

QY 430 AGTATCCCTGGAGAGTGGAGAGTTGATATTCAGTCTATTTTATCTTGTGATGTAATGGC 489
DB 51 AGTATTGCTTCTGATTCGAAAGATTCTGTTGATCTATTTGCTGTGGTGCAGAAAAGTT 110
QY 490 TTGCTTGTCCCTCAGAAAGTATTCGTTTGTGTTGGGATGAGACAA 535
DB 111 TTGTTGTCTCCCTCCGTTTTTGGCGCTTCCTTTTGAAGATTCTA 156

RESULT 2
US-09-107-532A-3543
; Sequence 3543; Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3543:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 552 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...552
; SEQUENCE DESCRIPTION: SEQ ID NO: 3543:
US-09-107-532A-3543
Query Match 5.8%; Score 38.8; DB 4; Length 552;
Best Local Similarity 60.4%; Pred. No. 0.019; Indels 0; Gaps 0;
Matches 64; Conservative 0; Mismatches 42;
```


Db 170 GGTTCCTGTTGGCTCCGTTGGTAAATAGTGGCTCTACAGGGTCTACAGGTGGCTCAATT 111

QY 451 GTTGATATTGAGTCTATTTTATCTTGATGTAATTCCTTTGCTTCTCCCTCA 504

Db 110 GCGCGTGTGATCTGTTGGCTTTTCTGGTGTACTTGGTCTGTTGGTACTTCA 57

RESULT 9

US-09-071-035-229/c

Sequence 229, Application US/09071035

Patent No. 6448043

GENERAL INFORMATION:

APPLICANT: Gil H. Choi

TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides

NUMBER OF SEQUENCES: 496

CORRESPONDENCE ADDRESS:

ADDRESSER: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/071.035

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: A. Anders Brookes

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB369P2

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 229:

SEQUENCE CHARACTERISTICS:

LENGTH: 834 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-09-071-035-229

Query Match 5.1%; Score 34; DB 4; Length 834;

Best Local Similarity 56.1%; Pred. No. 0.75; Indels 0; Gaps 0;

Matches 64; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 391 GTTCTGATTTGCTACTGATTTGAAAAATTTGTATCTACAGTATCCCTGGAGATGGAGA 450

Db 293 GGTTCGTTGGCTCCGTTGGTGTAAATAGTGGCTCTACAGGTCTACAGGTGGCTCAATT 234

QY 451 GTTGATATTGAGTCTATTTTATCTTGATGTAATTCCTTTGCTTCTCCCTCA 504

Db 233 GCGCGTGTGATCTGTTGGCTTTTCTGGTGTACTTGGTCTGTTGGTACTTCA 180

RESULT 10

US-08-045-806-1

Sequence 1, Application US/08045806

Patent No. 5378822

GENERAL INFORMATION:

APPLICANT: Bradfield, Christopher Alan

APPLICANT: Dolwick, Kristin Marie

APPLICANT: Poland, Alan

TITLE OF INVENTION: Ah Receptor cDNA and Method of

TITLE OF INVENTION: Determining Human Risks To Environmental Pollutants

NUMBER OF SEQUENCES: 23

ORGANISM: Mycobacterium tuberculosis

OTHER INFORMATION: H37RV

US-09-103-840A-1

Query Match 5.1%; Score 34.6; DB 3; Length 4411529;

Best Local Similarity 49.2%; Pred. No. 32;

Matches 91; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 162 ACATAGACGCTCCATTTGTCAGACAATCGGTGAGGGCGGAGCGCATTCGTCTG 221

Db 1379613 AGCAGAGCAACCCAGCATCCCGCTTGATCATCGGAGATGACCTTGTCAACGCCGC 1379554

QY 222 AAGGCACCTCGCTGCAAGAGCGACGCTGATTCACACCGTGGCTGACATGATCCAAAC 281

Db 1379553 CAGGCGCTGGCGGTCAACCTGCTGCTTAGCAGATCCTCGGGAGACCGCCCAAC 1379494

QY 282 AGACTGCAATTTGAGTACACCGAGCGGACGATCGGATCGCTCGGATGAAGATG 341

Db 1379493 CCGGCCAAAAGCAGCAGCAGACATAAAACGGCGTCTTCCAGACTCGCGATCACCAG 1379434

QY 342 ATGCC 346

Db 1379433 ATGCC 1379429

RESULT 8

US-09-071-035-231/c

Sequence 231, Application US/09071035

Patent No. 6448043

GENERAL INFORMATION:

APPLICANT: Gil H. Choi

TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides

NUMBER OF SEQUENCES: 496

CORRESPONDENCE ADDRESS:

ADDRESSER: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/071.035

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: A. Anders Brookes

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB369P2

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 231:

SEQUENCE CHARACTERISTICS:

LENGTH: 649 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-09-071-035-231

Query Match 5.1%; Score 34; DB 4; Length 649;

Best Local Similarity 56.1%; Pred. No. 0.66; Indels 0; Gaps 0;

Matches 64; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 391 GTTCTGATTTGCTACTGATTTGAAAAATTTGTATCTACAGTATCCCTGGAGATGGAGA 450

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Tilton, Fallon, Lungmus & Chestnut
;; STREET: 100 South Wacker Drive, Suite 960
;; CITY: Chicago
;; STATE: Illinois
;; COUNTRY: USA
;; ZIP: 60606-4002
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent in Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/045,806
;; FILING DATE: 19930408
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Pentress, Susan B.
;; REGISTRATION NUMBER: 31,327
;; REFERENCE/DOCKET NUMBER: NU-9207
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (312)-456-8000
;; TELEFAX: (312)-456-7776
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 3207 base pairs
;; TYPE: NUCLEIC ACID
;; STRANDEDNESS: double
;; TOPOLOGY: unknown
;; MOLECULE TYPE: cDNA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 1..2415
;; US-08-045-806-1

Query Match 5.0%; Score 33.6; DB 1; Length 3207;
Best Local Similarity 61.4%; Pred. No. 2;
Matches 54; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 390 TGTTCCTGATTCCTCACTGATGGAAATTTGTATCTACCAAGTATCCCTGGAGAGTGAG 449
DB 3004 TGTGTTGTTTCTGATTTTCTTTTCTATCTACCTGTAAACAATAGGCGTATGT 3063

QY 450 AGTTGATTCAGTCTATTTTATCTTGT 477
DB 3064 ATTTATATGAATAATTTTATCTTTT 3091

RESULT 11
US-08-366-051B-1
; Sequence 1, Application US/08366051B
; Patent No. 5650283
; GENERAL INFORMATION:
; APPLICANT: Bradfield, Christopher A.
; APPLICANT: Dolwick, Kristin M.
; APPLICANT: Carver, Lucy A.
; TITLE OF INVENTION: Ah Receptor cDNAs and Genetically
; TITLE OF INVENTION: Engineered Cells for Detecting Agonists to the Ah
; TITLE OF INVENTION: Receptor
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Tilton, Fallon, Lungmus & Chestnut
; STREET: 100 South Wacker Drive, Suite 960
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606-4002
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/366,051B
;; FILING DATE:
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Tilton, Timothy L.
;; REGISTRATION NUMBER: 16,926
;; REFERENCE/DOCKET NUMBER: NU-9207-CIP
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (312)-456-8000
;; TELEFAX: (312)-456-7776
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 3207 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: unknown
;; MOLECULE TYPE: cDNA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 1..2415
;; US-08-366-051B-1

Query Match 5.0%; Score 33.6; DB 1; Length 3207;
Best Local Similarity 61.4%; Pred. No. 2;
Matches 54; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 390 TGTTCCTGATTCCTCACTGATGGAAATTTGTATCTACCAAGTATCCCTGGAGAGTGAG 449
DB 3004 TGTGTTGTTTCTGATTTTCTTTTCTATCTACCTGTAAACAATAGGCGTATGT 3063

QY 450 AGTTGATTCAGTCTATTTTATCTTGT 477
DB 3064 ATTTATATGAATAATTTTATCTTTT 3091

RESULT 12
US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 650581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/557,884
; FILING DATE: 25-Apr-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,102
; FILING DATE: JUN-5-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB186P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:

1 APPLICANT: Myers, Lisa E
 2 APPLICANT: Schryvers, Anthony B
 3 APPLICANT: Harkness, Robin E
 4 APPLICANT: Loomore, Sheena M.
 5 APPLICANT: Du, Run-Pan
 6 APPLICANT: Yang, Yan-Ping
 7 APPLICANT: Klein, Michel H
 8 TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella
 9 NUMBER OF SEQUENCES: 60
 10 CORRESPONDENCE ADDRESS:

```

1 STATE: Ontario
2 COUNTRY: Canada
3 ZIP: M5G 1R7
4
5 COMPUTER READABLE FORM:
6 MEDIUM TYPE: Floppy disk
7

```

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: 05/09/059,584
 FILING DATE: 14-APR-1998
 CLASSIFICATION:

APPLICATION NUMBER: US 08/778,570
FILING DATE: 03-JAN-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:

NAME: Stewart, Michael I
REGISTRATION NUMBER: 24973
REFERENCE/DOCKET NUMBER: 1038-794
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155

TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 2145 base pairs

```

; ANNOT: 2110 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-059-584-48

```

Query Match. 4.8%; Score 32.6; DB 4; Length 2145;
Best Local Similarity 53.1%; Pred. NO. 3.4;
Matches 94; Conservative 0; Mismatches 79; Indels 4; Caps 1;

QY 477 TGATGTAATTGCCCTTTGCTTGTCTCCCTCAGAAAGTATTCTGTTGTTGTTGGGATGAGACAAAG 536

Db 663 TCGTTGGTTGTTTGTTCATTTGAATGGAAGTTTTCGCTTTTATCGATGATAACTTGG 604
Qy 537 TGGATTAAGAGTCTACTATATACACGATCATCTCTGTGTGAAGTTTGGCAGTTCTGCAG 596
Db 603 CTGAACCTCAGGCTTTTACGCACATT-----TCTTTGATTTTTTTGGCAAGTCTTCAT 548
Qy 597 TTCATGATCTGTAATTTGATGATGCTGGATTTCTACTATTATCAATCGTCATTAT 653
Db 547 CGGATATTTTATTTTGTTCATTTCTCTCTTAATTTCTTCCCAATGGCAATTT 491

RESULT 15

US-09-059-584-47/c
; Sequence 47, Application US/09059584
; Patent No. 6440701
; GENERAL INFORMATION:
; APPLICANT: Myers, Lisa E
; APPLICANT: Schryvers, Anthony B
; APPLICANT: Hartness, Robin E
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Du, Run-Pan
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/059,584
; FILING DATE: 14-APR-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/778,570
; FILING DATE: 03-JAN-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24973
; REFERENCE/DOCKET NUMBER: 1038-794
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 47:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2287 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-059-584-47

Query Match 4.8%; Score 32.6; DB 4; Length 2287;
Best Local Similarity 53.1%; Pred. No. 3.5;
Matches 94; Conservative 0; Mismatches 79; Indels 4; Gaps 1;
Qy 477 TGATGTAATTCGCTTTGCTTCCTCAGAGATATTCGTTTGTGGGATGAGACAAG 536
Db 805 TCGTTTGGTTGTTTGTTCATTTGAATGGAAGTTTTCGCTTTTATCGATGATAACTTGG 746
Qy 537 TGGATTAAGAGTCTACTATATACACGATCATCTCTGTGTGAAGTTTGGCAGTTCTGCAG 596
Db 745 CTGAACCTCAGGCTTTTACGCACATT-----TCTTTGATTTTTTGGCAAGTCTTCAT 690

Qy 597 TTTCATGATCTCTAATTTGATGATGCTGGATTTCTACTATTATCAATCGTCATTAT 653
Db 689 CGGATATTTTATTTTGTTCATTTTCTCTCTTAATTTCTTCCCAATGGCAATTT 633
Search completed: November 14, 2003, 09:21:39
Job time : 81 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2003, 06:42:02 ; Search time 223 Seconds

(without alignments)
8146.735 Million cell updates/sec

Title: US-09-868-025-1

Perfect score: 673

Sequence: 1 ttatccttgcctgctcgat.....actgtgtgtaaaaaaa 673

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_19Jun03.*
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	673	100.0	673	22 AAH19867	Rice AGT-SAL 11 po
2	121	18.0	888	24 ABN98736	Arabidopsis thalia
3	121	18.0	1247	21 AAC54660	Arabidopsis thalia
4	121	18.0	1252	21 AAC54662	Arabidopsis thalia
5	121	18.0	1253	21 AAC53587	Arabidopsis thalia
6	110.6	16.4	1255	21 AAC54330	Arabidopsis thalia
7	105.2	15.6	1035	24 ABZ14447	Arabidopsis thalia
8	105.2	15.6	1346	21 AAC54760	Arabidopsis thalia

9	- 105.2	15.6	1349	21 AAC35078	Arabidopsis thalia
10	44.8	6.7	490	21 AAC37164	Arabidopsis thalia
11	43.8	6.5	7624	24 ABL34112	Human immune syste
12	42	6.2	7634	24 ABL34130	Human immune syste
13	39.6	5.9	7449	24 ABL32276	Human immune syste
14	39.4	5.9	19459	24 ABL70527	Chemically treated
15	39.4	5.9	19459	24 ABK31212	Signal transductio
16	39	5.8	10279	24 ABL92276	Chemically treated
17	39	5.8	10279	24 ABL33590	Human immune syste
18	39	5.8	10279	24 ABL33590	Chemically treated
19	39	5.8	10279	24 ABL33590	Human immune syste
20	38.6	5.7	8479	23 ABZ11112	Human nervous syst
21	38	5.6	14491	25 ABZ10061	Drosophila melanog
22	38	5.6	14491	25 ABZ10061	Haematopoietic cel
23	38	5.6	17491	24 ABL34574	Haematopoietic cel
24	37	5.5	2644	22 AAF77695	Human Trafa4 bindin
25	36.8	5.5	7784	24 ABL34424	Human immune syste
26	36.8	5.5	7784	24 ABL33687	Human immune syste
27	36.8	5.5	14712	24 ABN80249	Human chemically m
28	36.4	5.4	5518	25 ABZ10115	Haematopoietic cel
29	36.4	5.4	684707	24 ABQ67196	Haematopoietic cel
30	36.4	5.4	3011208	24 ABQ69245	Listeria innocua c
31	36	5.3	991	22 AAI90690	Listeria innocua c
32	36	5.3	6051	22 AAS46604	Human polynucleoti
33	36	5.3	6051	24 ABL33851	Tumour suppressor
34	35.8	5.3	3060	22 AHS4131	Human immune syste
35	35.8	5.3	3543	22 AHS4129	S. epidermidis gen
36	35.8	5.3	5228	24 ABL33655	S. epidermidis gen
37	35.8	5.3	11091	24 ABN92780	Human immune syste
38	35.6	5.3	2606	23 ABL02779	Staphylococcus epi
39	35.6	5.3	3545	23 ABL02780	Drosophila melanog
40	35.6	5.3	5215	23 ABL02778	Drosophila melanog
41	35.6	5.3	9021	22 AAS46325	Tumour suppressor
42	35.6	5.3	10552	24 ABL70388	Chemically treated
43	35.6	5.3	10552	24 AAS61340	Human gene regulat
44	35.6	5.3	10552	24 ABK31427	Signal transductio
45	35.6	5.3	12639	24 ABN80107	Human chemically m

ALIGNMENTS

RESULT 1

AAH19867

ID AAH19867 standard; DNA; 673 BP.

XX AAH19867;

AC AAH19867;

DT 03-AUG-2001 (first entry)

XX Rice AGT-SAL 11 polynucleotide sequence SEQ ID NO:1.

DE Rice AGT-SAL 11; salt tolerance; plant; cytostatic; antiHIV;

XX Rice AGT-SAL 11; salt tolerance; plant; cytostatic; antiHIV;

KW proteinase inhibitor; cancer; human immunodeficiency virus;

KM HIV infection; animal disorder; food processing; enzyme industry;

XX biological preservative; ds.

XX Oryza sativa.

OS Oryza sativa.

PN WO200130990-A2.

XX WO200130990-A2.

PD 03-MAY-2001.

XX 11-OCT-2000; 2000WO-IN00099.

PF 13-OCT-1999; 99IN-0000997.

XX (AVES-) AVESTHAGEN GRAINE TECHNOLOGIES PVT LTD.

PR Patell VM, Antony CM, Chandran D, Madurappa A;

XX WPI; 2001-308632/32.

XX P-PSDB; AAB75128.

DR

XX Nucleotide sequence encoding an AGR-SAL 11 polypeptide similar to
PT Bowman Birk II type proteinase inhibitors is useful to confer salt
PT resistance to plants -
XX
PS Claim 1; Page 17; 22pp; English.
XX
CC The present sequence encodes the rice AGR-SAL 11 protein. The AGR-SAL 11
CC protein can be used to confer salt tolerance to plants and other
CC organisms. The AGR-SAL 11 gene was isolated from salt-stressed rice.
CC Also described are: (i) a transgenic plant comprising a recombinant
CC expression cassette comprising a plant promoter operably linked to N1;
CC and (ii) conferring salt tolerance on a plant, comprising introducing
CC the above expression cassette. The AGR-SAL 11 protein is a proteinase
CC inhibitor. The proteinase inhibitor may be used to confer stress
CC tolerance to many plants and organisms including cotton, maize, rice,
CC soybean, sugar beet, wheat, fruit, vegetables and vines, particularly
CC biotic bacterial, fungal and pest stresses. Proteinase inhibitors are
CC also useful in the treatment of cancer, human immunodeficiency virus
CC (HIV) infection and other animal disorders. The gene may also be useful
CC in food processing and enzyme industries as an inhibitor of proteinase
CC activity and as a biological preservative.
XX
SQ Sequence 673 BP; 163 A; 145 C; 170 G; 195 T; 0 other;
Query Match 100.0%; Score 673; DB 22; Length 673;
Best Local Similarity 100.0%; Pred. No. 3.7e-202;
Matches 673; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTTACCTTGCTGCTCGATGGAGCAAACTTCATCTGGGGTGGCGTGACACCA 60
DB 1 TTTACCTTGCTGCTCGATGGAGCAAACTTCATCTGGGGTGGCGTGACACCA 60
QY 61 AGAAATTCCTCCCTCAGTGGTTTGCAGCTGTCATCGCGTGTTCATTGGAATGC 120
DB 61 AGAAATTCCTCCCTCAGTGGTTTGCAGCTGTCATCGCGTGTTCATTGGAATGC 120
QY 121 TGAGGAATCTGCAACATGCCAGACTGCCATGCCATTCACATAGCAGCTCCATTG 180
DB 121 TGAGGAATCTGCAACATGCCAGACTGCCATTCACATAGCAGCTCCATTG 180
QY 181 TTGTCAGCAATCGGTCGAGGCGGAGCGCATTCGTCTGAAGGCACTGGCTGCAAGA 240
DB 181 TTGTCAGCAATCGGTCGAGGCGGAGCGCATTCGTCTGAAGGCACTGGCTGCAAGA 240
QY 241 GCGACGCTGATTCCACCACCGTGGCTGACATGTATCCAAACAGACTGCAAAATGCA 300
DB 241 GCGACGCTGATTCCACCACCGTGGCTGACATGTATCCAAACAGACTGCAAAATGCA 300
QY 301 ACACGAGGCGAAGCGATGGATCCGCTCGCGATGAAGATGATCGCGGCGGCTTCTG 360
DB 301 ACACGAGGCGAAGCGATGGATCCGCTCGCGATGAAGATGATCGCGGCGGCTTCTG 360
QY 361 GTGTCGCTGCTGCTCCACACCAAGCATGTGTTTCTGATGCTCAGTGTGAAATTT 420
DB 361 GTGTCGCTGCTGCTCCACACCAAGCATGTGTTTCTGATGCTCAGTGTGAAATTT 420
QY 421 GTATCTACCATGATCCCTCGAGAGTGGAGATGATATGAGTCTATTTATCTTGAT 480
DB 421 GTATCTACCATGATCCCTCGAGAGTGGAGATGATATGAGTCTATTTATCTTGAT 480
QY 481 GTAATTCCTTGTCTGCTCCCTCAAGATATGCTGTTTGTGGATGAGCAAGTGA 540
DB 481 GTAATTCCTTGTCTGCTCCCTCAAGATATGCTGTTTGTGGATGAGCAAGTGA 540
QY 541 ATAAGAGTCTACTATATACAGATCATCTGTTTAAAGTTTCCAGTTCTCGAGTTCA 600
DB 541 ATAAGAGTCTACTATATACAGATCATCTGTTTAAAGTTTCCAGTTCTCGAGTTCA 600
QY 601 TGTATCTGTAATTTGATGATGCTGGATTTCTACTATTTATCAATGCTATTACTGTT 660
DB 601 TGTATCTGTAATTTGATGATGCTGGATTTCTACTATTTATCAATGCTATTACTGTT 660

QY 661 GTAAAAA 673
DB 661 GTAAAAA 673
RESULT 2
ABN98736/C
ID ABN98736 standard; DNA; 888 BP.
XX
AC ABN98736;
XX
DT 01-AUG-2002 (first entry)
XX
DE Arabidopsis thaliana expressed polynucleotide SEQ ID NO 504.
XX
KW Arabidopsis thaliana; plant; insecticide; fungicide; transgenic; stress;
KW disease; crop; thale cress; tolerance factor; insect; pathogen;
KW nutrition; ds.
XX
OS Arabidopsis thaliana.
XX
PN US2002023281-A1.
XX
PD 21-PEB-2002.
XX
PF 26-JAN-2001; 2001US-0770445.
XX
PR 27-JAN-2000; 2000US-178472P.
XX
PA (GORL/) GORLACH J.
PA (ANY/) AN Y.
PA (HAMI/) HAMILTON C M.
PA (PRIC/) PRICE J L.
PA (RAIN/) RAINES T M.
PA (YUY/) YU Y.
PA (RAME/) RAMEAKA J G.
PA (PAGE/) PAGE A.
PA (MATH/) MATHW A V.
PA (LEDF/) LEDFORD B L.
PA (WORS/) WORSNER J P.
PA (HAAS/) HAAS W D.
PA (GARC/) GARCIA C A.
PA (KRIC/) KRICKER M.
PA (SLAT/) SLATER T.
PA (DAVI/) DAVIS K R.
PA (ALLE/) ALLEN K.
PA (HOFF/) HOFFMAN N.
PA (HURB/) HURBAN P.
XX
PI Gorlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y;
PI Rameaka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;
PI Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;
PI Hurban P;
XX
DR WPI; 2002-400781/43.
XX
PT New Arabidopsis thaliana nucleic acid for identifying homologous genes,
PT producing compositions that modulate the expression or function of its
PT encoded protein, and mapping functional regions of protein -
XX
PS Claim 1; SEQ ID NO 504; 49pp + Sequence Listing; English.
XX
CC The invention relates to an Arabidopsis thaliana nucleic acid (I)
CC comprising a sequence capable of hybridising under stringent conditions
CC to a sequence selected from any one of 999 sequences (ABN98233-ABN99233),
CC given in the specification or its fragment. A polypeptide (II) encoded by
CC (I), a transgenic plant (III) comprising an exogenous nucleic acid or a
CC genetically modified cell (IV) comprising an exogenous nucleic acid, is
CC useful for screening a candidate agent for its biological effect. (I) is
CC useful in identifying homologous or related genes, in producing
CC compositions that modulate the expression or function of its encoded
CC protein, mapping functional regions of the protein and in studying
CC associated physiological pathways. (I) is also useful for the genetic

PR	12-JUL-1999;	99US-0142977.	PR	29-SEP-1999;	99US-0156596.
PR	13-JUL-1999;	99US-0143542.	PR	04-OCT-1999;	99US-0157117.
PR	14-JUL-1999;	99US-0143624.	PR	05-OCT-1999;	99US-0157753.
PR	15-JUL-1999;	99US-0144005.	PR	06-OCT-1999;	99US-0157865.
PR	16-JUL-1999;	99US-0144085.	PR	07-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144086.	PR	08-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144325.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144333.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159637.
PR	20-JUL-1999;	99US-0144884.	PR	14-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0144814.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160767.
PR	21-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160981.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160989.
PR	26-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161404.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161406.
PR	27-JUL-1999;	99US-0145919.	PR	26-OCT-1999;	99US-0161359.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146388.	PR	28-OCT-1999;	99US-0161920.
PR	02-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161992.
PR	03-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161993.
PR	04-AUG-1999;	99US-0147204.	PR	29-OCT-1999;	99US-0162142.
PR	04-AUG-1999;	99US-0147302.			
PR	05-AUG-1999;	99US-0147192.			
PR	05-AUG-1999;	99US-0147260.			
PR	06-AUG-1999;	99US-0147303.			
PR	06-AUG-1999;	99US-0147416.			
PR	09-AUG-1999;	99US-0147493.			
PR	09-AUG-1999;	99US-0147935.			
PR	10-AUG-1999;	99US-0148171.			
PR	11-AUG-1999;	99US-0148319.			
PR	12-AUG-1999;	99US-0148341.			
PR	13-AUG-1999;	99US-0148565.			
PR	13-AUG-1999;	99US-0148684.			
PR	16-AUG-1999;	99US-0149368.			
PR	17-AUG-1999;	99US-0149175.			
PR	18-AUG-1999;	99US-0149428.			
PR	20-AUG-1999;	99US-0149722.			
PR	20-AUG-1999;	99US-0149723.			
PR	20-AUG-1999;	99US-0149923.			
PR	23-AUG-1999;	99US-0149902.			
PR	23-AUG-1999;	99US-0149930.			
PR	25-AUG-1999;	99US-0150566.			
PR	26-AUG-1999;	99US-0150884.			
PR	27-AUG-1999;	99US-0151065.			
PR	27-AUG-1999;	99US-0151066.			
PR	27-AUG-1999;	99US-0151080.			
PR	30-AUG-1999;	99US-0151303.			
PR	31-AUG-1999;	99US-0151438.			
PR	01-SEP-1999;	99US-0151930.			
PR	07-SEP-1999;	99US-0152363.			
PR	10-SEP-1999;	99US-0153070.			
PR	13-SEP-1999;	99US-0153758.			
PR	15-SEP-1999;	99US-0154018.			
PR	16-SEP-1999;	99US-0154039.			
PR	20-SEP-1999;	99US-0154779.			
PR	22-SEP-1999;	99US-0155139.			
PR	23-SEP-1999;	99US-0155485.			
PR	24-SEP-1999;	99US-0155659.			
PR	28-SEP-1999;	99US-0156458.			

Query Match 18.0%; Score 121; DB 21; Length 1247;
 Best Local Similarity 68.9%; Pred. No. 2e-27;
 Matches 166; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY	27	AAATCCATCTTGGGGTGGGTGGGAGACACACACAGAAATCTCCCTCAGTGGTTGCA	86
Db	716	AATGGCCCTTAGGTGTATGGAGAGACACACAAAGAGTTTCGGCTCATGGTTTGG	775
QY	87	GCTGTCCATCGCGCTGTTCCATTCATTGGAATGCTGAGGAATCTGTCAACATGCCCAAG	146
Db	776	GCTGTCCATCGCTGCTGTGCTTTTATAGCAATGCTTAGGAATCTGTCTGATGCCCAA	835
QY	147	ACTGCCATGCGATTCCACCATAGAGCTCCATTTGTTGGTCAGACATCGGTCGAGGGCG	206
Db	836	ACAGCCATGGCTTTGACCAITGGAGCTTCGATCTTGGGACAGTGATGGGTCAGAGCT	895
QY	207	GAGCGCATTCGTCTGAAGGCACTGGCTGCAAGAGCGAGCTGATTCCACCCGTTGGCT	266
Db	896	GAGCGTTACCGTCTCAAGCAGTAGCTGAGAAATGGTTCTCTGTAACCGCTATGGTTCT	955
QY	267	G 267	
Db	956	G 956	

RESULT 4
 AAC54662
 ID AAC54662 standard; DNA; 1252 BP.
 XX AAC54662;
 AC
 XX
 DT 18-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana DNA fragment SEQ ID NO: 78642.
 XX Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway;

KW metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 08-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

XX 19-APR-1999; 99US-0130077.

XX 21-APR-1999; 99US-0130449.

XX 23-APR-1999; 99US-0130510.

XX 23-APR-1999; 99US-0130891.

XX 28-APR-1999; 99US-0131449.

XX 30-APR-1999; 99US-0132048.

XX 30-APR-1999; 99US-0132407.

XX 04-MAY-1999; 99US-0132484.

XX 05-MAY-1999; 99US-0132485.

XX 06-MAY-1999; 99US-0132486.

XX 06-MAY-1999; 99US-0132487.

XX 07-MAY-1999; 99US-0132863.

XX 11-MAY-1999; 99US-0134256.

XX 14-MAY-1999; 99US-0134218.

XX 14-MAY-1999; 99US-0134219.

XX 14-MAY-1999; 99US-0134221.

XX 14-MAY-1999; 99US-0134370.

XX 18-MAY-1999; 99US-0134768.

XX 19-MAY-1999; 99US-0134941.

XX 20-MAY-1999; 99US-0135124.

XX 21-MAY-1999; 99US-0135353.

XX 24-MAY-1999; 99US-0135629.

XX 25-MAY-1999; 99US-0136021.

XX 27-MAY-1999; 99US-0136392.

XX 28-MAY-1999; 99US-0136782.

XX 01-JUN-1999; 99US-0137222.

XX 03-JUN-1999; 99US-0137528.

XX 04-JUN-1999; 99US-0137502.

XX 07-JUN-1999; 99US-0137724.

XX 08-JUN-1999; 99US-0138094.

XX 10-JUN-1999; 99US-0138540.

XX 14-JUN-1999; 99US-0138847.

XX 16-JUN-1999; 99US-0139119.

XX 16-JUN-1999; 99US-0139452.

XX 17-JUN-1999; 99US-0139453.

XX 18-JUN-1999; 99US-0139454.

XX 18-JUN-1999; 99US-0139455.

XX 18-JUN-1999; 99US-0139456.

XX 18-JUN-1999; 99US-0139457.

XX 18-JUN-1999; 99US-0139458.

XX 18-JUN-1999; 99US-0139459.

XX 18-JUN-1999; 99US-0139460.

XX 18-JUN-1999; 99US-0139461.

XX 18-JUN-1999; 99US-0139462.

XX 18-JUN-1999; 99US-0139463.

XX 18-JUN-1999; 99US-0139750.

XX 18-JUN-1999; 99US-0139763.

XX 21-JUN-1999; 99US-0139817.

XX 22-JUN-1999; 99US-0139899.

XX 23-JUN-1999; 99US-0140353.

XX 23-JUN-1999; 99US-0140354.

XX 24-JUN-1999; 99US-0140595.

XX 28-JUN-1999; 99US-0140823.

XX 29-JUN-1999; 99US-0140991.

XX 30-JUN-1999; 99US-0141287.

XX 01-JUL-1999; 99US-0141842.

XX 01-JUL-1999; 99US-0142154.

XX 02-JUL-1999; 99US-0142055.

XX 06-JUL-1999; 99US-0142390.

XX 08-JUL-1999; 99US-0142803.

XX 09-JUL-1999; 99US-0142920.

XX 12-JUL-1999; 99US-0142977.

XX 13-JUL-1999; 99US-0143542.

XX 14-JUL-1999; 99US-0143624.

XX 15-JUL-1999; 99US-0144005.

XX 16-JUL-1999; 99US-0144085.

XX 16-JUL-1999; 99US-0144086.

XX 19-JUL-1999; 99US-0144325.

XX 19-JUL-1999; 99US-0144331.

XX 19-JUL-1999; 99US-0144332.

XX 19-JUL-1999; 99US-0144333.

XX 19-JUL-1999; 99US-0144334.

XX 19-JUL-1999; 99US-0144335.

XX 20-JUL-1999; 99US-0144352.

XX 20-JUL-1999; 99US-0144632.

XX 20-JUL-1999; 99US-0144884.

XX 21-JUL-1999; 99US-0144814.

XX 21-JUL-1999; 99US-0145086.

XX 21-JUL-1999; 99US-0145088.

XX 22-JUL-1999; 99US-0145085.

XX 22-JUL-1999; 99US-0145087.

XX 22-JUL-1999; 99US-0145089.

XX 22-JUL-1999; 99US-0145192.

XX 23-JUL-1999; 99US-0145145.

XX 23-JUL-1999; 99US-0145218.

XX 23-JUL-1999; 99US-0145224.

XX 26-JUL-1999; 99US-0145276.

XX 27-JUL-1999; 99US-0145913.

XX 27-JUL-1999; 99US-0145918.

XX 27-JUL-1999; 99US-0145919.

XX 28-JUL-1999; 99US-0145951.

XX 02-AUG-1999; 99US-0146386.

XX 02-AUG-1999; 99US-0146388.

XX 02-AUG-1999; 99US-0146389.

XX 03-AUG-1999; 99US-0147038.

XX 04-AUG-1999; 99US-0147204.

XX 04-AUG-1999; 99US-0147302.

XX 05-AUG-1999; 99US-0147192.

XX 05-AUG-1999; 99US-0147260.

XX 06-AUG-1999; 99US-0147303.

XX 06-AUG-1999; 99US-0147416.

XX 09-AUG-1999; 99US-0147493.

XX 09-AUG-1999; 99US-0147935.

XX 10-AUG-1999; 99US-0148171.

XX 11-AUG-1999; 99US-0148319.

XX 12-AUG-1999; 99US-0148341.

XX 13-AUG-1999; 99US-0148565.

XX 13-AUG-1999; 99US-0148684.

XX 16-AUG-1999; 99US-0149368.

XX 17-AUG-1999; 99US-0149175.

XX 18-AUG-1999; 99US-0149426.

XX 20-AUG-1999; 99US-0149722.

XX 20-AUG-1999; 99US-0149723.

XX 23-AUG-1999; 99US-0149929.

XX 23-AUG-1999; 99US-0149902.

XX 23-AUG-1999; 99US-0149930.

XX 25-AUG-1999; 99US-0150566.

XX 26-AUG-1999; 99US-0150884.

XX 27-AUG-1999; 99US-0151065.

XX 27-AUG-1999; 99US-0151066.

XX 30-AUG-1999; 99US-0151080.

XX 31-AUG-1999; 99US-0151303.

XX 31-AUG-1999; 99US-0151438.

XX 01-SEP-1999; 99US-0151910.

XX 07-SEP-1999; 99US-0152383.

PR	10-SEP-1999	99US-0153070
PR	11-SEP-1999	99US-0153078
PR	13-SEP-1999	99US-0153758
PR	15-SEP-1999	99US-0154036
PR	16-SEP-1999	99US-0154039
PR	20-SEP-1999	99US-0154779
PR	22-SEP-1999	99US-0155133
PR	23-SEP-1999	99US-0155139
PR	24-SEP-1999	99US-0155486
PR	24-SEP-1999	99US-0155659
PR	28-SEP-1999	99US-0156458
PR	28-SEP-1999	99US-0156596
PR	04-OCT-1999	99US-0157117
PR	05-OCT-1999	99US-0157753
PR	06-OCT-1999	99US-0157865
PR	06-OCT-1999	99US-0158065
PR	08-OCT-1999	99US-0158232
PR	12-OCT-1999	99US-0158369
PR	13-OCT-1999	99US-0159293
PR	13-OCT-1999	99US-0159294
PR	14-OCT-1999	99US-0159295
PR	14-OCT-1999	99US-0159329
PR	14-OCT-1999	99US-0159330
PR	14-OCT-1999	99US-0159331
PR	14-OCT-1999	99US-0159637
PR	14-OCT-1999	99US-0159638
PR	18-OCT-1999	99US-0159584
PR	21-OCT-1999	99US-0160741
PR	21-OCT-1999	99US-0160767
PR	21-OCT-1999	99US-0160768
PR	21-OCT-1999	99US-0160770
PR	21-OCT-1999	99US-0160814
PR	21-OCT-1999	99US-0160815
PR	22-OCT-1999	99US-0160980
PR	22-OCT-1999	99US-0160981
PR	22-OCT-1999	99US-0160989
PR	23-OCT-1999	99US-0161404
PR	23-OCT-1999	99US-0161405
PR	23-OCT-1999	99US-0161406
PR	23-OCT-1999	99US-0161359
PR	26-OCT-1999	99US-0161360
PR	26-OCT-1999	99US-0161361
PR	28-OCT-1999	99US-0161920
PR	28-OCT-1999	99US-0161992
PR	28-OCT-1999	99US-0161993
PR	28-OCT-1999	99US-0162141
PR	28-OCT-1999	99US-0162142

Query Match	18.0%;	Score 121;	DB 21;	Length 1252;	
Best Local Similarity	68.9%;	Pred. No. 2e-27;			
Matches 166;	Conservative 0;	Mismatches 75;	Indels 0;	Gaps 0	
Qy	27	AAATCCCATCTTTGGGGTGTGGCGTGAGCACACCAAGAAATTCCTCCCTCAGTGGTTCGCA	86		
Db	721	AATGTGCCTCTAGGTGTATGGAGAAACACACAAGAAAGTTTCGCGTCGATGGTTTTGG	780		
Qy	87	GCGTGCCATCGCGCTGTTCCATTCATTTGGATGCTGAGGAAATCTGTCACATGCCACAG	146		
Db	781	GTGTGCCATGCTGCTGTGCCTTTTATGCAATGTTAGGAAATCTGTTCTGTATGCCCAA	840		
Qy	147	ACTGCCATGGCAATTCACCATAGCAGCTCCATTTGCTCAGACAAATCGGGTCGAGGGCG	206		
Db	841	ACAGCCATGGCTTTGACCATTTGGAGCTTCGATCTTGGGACAGGTGATGGGTCAAGAGCT	900		
Qy	207	GAGCGCATTCGTCTGAAGCACTGGCTGCAAGAGCGACGCTGATTCACACACCGTGGCT	266		
Db	901	GAGCGTTACGGTCTCAAGCAGTACTGAGAAATGTTCTCTGTACCGGTATGGTTTCT	960		
Qy	267	G 267			
Db	961	G 961			

RESULT 5
AAC53587
ID AAC53587 standard; DNA; 1253 BP.

XX	AAC53587;	
XX	18-OCT-2000 (first entry)	
XX	Arabidopsis thaliana DNA fragment SEQ ID NO: 75020.	
XX	Hybridisation assay; genetic mapping; gene expression control;	
XX	protein identification; signal transduction pathway;	
XX	metabolic pathway; promoter; termination sequence; ss.	
XX	Arabidopsis thaliana.	
XX	EP1033405-A2.	
XX	06-SEP-2000.	
XX	25-FEB-2000; 2000EP-0301439.	
XX	25-FEB-1999;	99US-0121825.
XX	05-MAR-1999;	99US-0123180.
XX	09-MAR-1999;	99US-0123548.
XX	23-MAR-1999;	99US-0125788.
XX	25-MAR-1999;	99US-0126264.
XX	29-MAR-1999;	99US-0126785.
XX	01-APR-1999;	99US-0127462.
XX	06-APR-1999;	99US-0128234.
XX	08-APR-1999;	99US-0128714.
XX	16-APR-1999;	99US-0129845.
XX	19-APR-1999;	99US-0130077.
XX	21-APR-1999;	99US-0130449.
XX	23-APR-1999;	99US-0130510.
XX	23-APR-1999;	99US-0130891.
XX	28-APR-1999;	99US-0131449.
XX	30-APR-1999;	99US-0132048.
XX	30-APR-1999;	99US-0132407.
XX	04-MAY-1999;	99US-0132484.
XX	05-MAY-1999;	99US-0132485.
XX	06-MAY-1999;	99US-0132486.
XX	07-MAY-1999;	99US-0132487.
XX	11-MAY-1999;	99US-0132853.
XX	14-MAY-1999;	99US-0134256.
XX	14-MAY-1999;	99US-0134218.
XX	14-MAY-1999;	99US-0134219.
XX	14-MAY-1999;	99US-0134221.
XX	18-MAY-1999;	99US-0134370.
XX	19-MAY-1999;	99US-0134768.
XX	20-MAY-1999;	99US-0134941.
XX	21-MAY-1999;	99US-0135124.
XX	24-MAY-1999;	99US-0135253.
XX	25-MAY-1999;	99US-0135629.
XX	27-MAY-1999;	99US-0136021.
XX	28-MAY-1999;	99US-0136392.
XX	01-JUN-1999;	99US-0136782.
XX	03-JUN-1999;	99US-0137221.
XX	04-JUN-1999;	99US-0137528.
XX	07-JUN-1999;	99US-0137502.
XX	08-JUN-1999;	99US-0137724.
XX	10-JUN-1999;	99US-0138094.
XX	10-JUN-1999;	99US-0138540.
XX	10-JUN-1999;	99US-0138847.
XX	14-JUN-1999;	99US-0139119.
XX	16-JUN-1999;	99US-0139452.
XX	16-JUN-1999;	99US-0139453.
XX	17-JUN-1999;	99US-0139492.
XX	18-JUN-1999;	99US-0139454.
XX	18-JUN-1999;	99US-0139455.
XX	18-JUN-1999;	99US-0139456.
XX	18-JUN-1999;	99US-0139457.
XX	18-JUN-1999;	99US-0139458.
XX	18-JUN-1999;	99US-0139459.
XX	18-JUN-1999;	99US-0139460.
XX	18-JUN-1999;	99US-0139461.

```
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139889.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145132.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145921.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147284.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147132.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.

PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 28-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 15-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 08-OCT-1999; 99US-0157865.
PR 08-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 18.0%; Score 121; DB 21; Length 1253;
Best Local Similarity 68.9%; Pred. No. 2e-27;
Matches 166; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 27 AAATCCATCTTGGGCTGGCGTAGACACACAAAGAAATCTCCCTCAGTGTTTGA 86
DB 722 AATGTGCTCTAGGTGTATGAGAGAACACACAAAGAGTTCGCTTCATGGTTTG 781
QY 87 GCTGTCCATGCGCTGTCTCCATTCATTGGAATGCTGAGGAATCTGTCAACATGCCAAG 146
DB 782 GCATGCCATGCTGCTGTGCTTTTATAGCAATGCTTAGGAATCTGTGCTGATGCCAAA 841
QY 147 ACTGCCATGGCAATCACCATAGCAGCTCCATTTGTTGTCAGCAATCGGGTCGAGGGG 206
DB 842 ACAGCCATGGCTTTCACCATTCGAGCTTCGATCTTTGGGACAGGTGATTTGGGTCAAGAGCT 901
QY 207 GAGCGCATTCGTCTGAAGGCACTGCTCAAGAGGAGCGCTGATTCCACCAACCGTGGCT 266
DB 902 GAGCGTACCGTCTCAAGCAGTAGCTGAGAAATGTTCTCTGAACCGCTATGTTTCT 961
```

Qy 267 G 267
Db 962 G 962

RESULT 6
AAC54330
ID AAC54330 standard; DNA; 1255 BP.
XX
AC AAC54330;
DT 18-OCT-2000 (first entry)
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 77495.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136592.
PR 28-MAY-1999; 99US-0136762.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145275.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147491.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.

785 GCTGTCATGCTGCTGCTGCTTTTATAGCAATGCTTAGGAAATCTGTTCTGATGCCCAA 844

147 ACTGCGATGGCAATCACCATAGCAGCTCCATGTTGGTCACACATCGGTCGAGGGC 206

845 ACAGCCATGCTTGGACCATGAGCTTCGATCTTGGGACAGGTGAT-GGGTCAAGACT 903

207 GAGCGCATTCGCTGAAGGCACTGCTGCAAGAGCGAGCTGATTCACACCGCTGGCT 266

904 GAGCGTACCGTCTCAAGCAGTAGCTGAGAAATGGTTCTCTGTAACCGCTATGTTCT 963

267 G 267

964 G 964

RESULT 7

ABZ14447

ID ABZ14447 standard; DNA; 1035 BP.

XX ABZ14447;

AC ABZ14447;

DT 21-JAN-2003 (First entry)

XX Arabidopsis thaliana stress regulated gene SEQ ID NO 2252.

DE Arabidopsis thaliana; plant; gene; stress; transgenic; ds.

XX Arabidopsis thaliana.

OS Arabidopsis thaliana.

XX MO200216655-A2.

PN 28-FEB-2002.

PD 24-AUG-2001; 2001WO-US26685.

XX 24-AUG-2000; 2000US-227866P.

PR 26-JAN-2001; 2001US-264647P.

XX 22-JUN-2001; 2001US-300111P.

XX (SCRI) SCRIPTS RES INST.

PA (SYGN) SYNGENTA PARTICIPATIONS AG.

XX Harper JF, Kreps J, Wang X, Zhu T;

PI MPI; 2002-304127/34.

XX Identifying a stress condition to which a plant cell has been exposed and producing plants with increased tolerance to these abiotic stresses

PT

PS Claim 144; SEQ ID NO 2252; 577bp + Sequence Listing; English.

XX The invention relates to identifying a stress condition to which a plant cell has been exposed, comprising:

CC (a) contacting nucleic acid representative of expressed polynucleotides in the plant cell with an array or probes representative of the plant cell genome; and

CC (b) detecting a profile of expressed polynucleotides in the plant cell characteristic of a stress response. The method is useful in the production of transgenic plants, cells and seeds and in producing plants with increased tolerance to abiotic stress. The present sequence is that of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used in methods of the invention.

CC Note: The sequence data for this patent is not represented in the printed CC specification but is based on sequence information supplied to the printed CC the European Patent Office.

XX SQ Sequence 1035 BP; 282 A; 235 C; 247 G; 271 T; 0 other;

Query Match 15.6%; Score 105.2; DB 24; Length 1035;

Best Local Similarity 68.2%; Pred. No. 1.8e-22;

Matches 146; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

99US-0148684.

99US-0149368.

99US-0149175.

99US-0149426.

99US-0149722.

99US-0149723.

99US-0149929.

99US-0149902.

99US-0149930.

99US-0150566.

99US-0150884.

99US-0151085.

99US-0151086.

99US-0151080.

99US-0151303.

99US-0151438.

99US-0151930.

99US-0152363.

99US-0153070.

99US-0153758.

99US-0154018.

99US-0154039.

99US-0154779.

99US-0155139.

99US-0155486.

99US-0155659.

99US-0156458.

99US-0156596.

99US-0157117.

99US-0157753.

99US-0157865.

99US-0158029.

99US-0158232.

99US-0158369.

99US-0159293.

99US-0159294.

99US-0159295.

99US-0159329.

99US-0159330.

99US-0159331.

99US-0159637.

99US-0159638.

99US-0159584.

99US-0160741.

99US-0160767.

99US-0160768.

99US-0160770.

99US-0160814.

99US-0160815.

99US-0160980.

99US-0160981.

99US-0160989.

99US-0161404.

99US-0161405.

99US-0161406.

99US-0161359.

99US-0161360.

99US-0161361.

99US-0161920.

99US-0161992.

99US-0161993.

99US-0162142.

Query Match 16.4%; Score 110.6; DB 21; Length 1255;

Best Local Similarity 68.4%; Pred. No. 3.8e-24;

Matches 166; Conservative 0; Mismatches 74; Indels 1; Gaps 1;

27 AACTCCATCTGGGGTGGGTCGAGCACCAAGAAATCTCCCTCAGTGGTTGCA 86

725 AATGTCCTCTAGCTGTATGAGAGACACACAAAGAGTTTTCGCTGCAATGTTTG 784

87 GCTGTCCATGCGCTGTTTCCATTCAATGGAATGCTGAGAAATCTGCAACATGCCCAAG 146

QY 27 AAACCTCCATCTTGGGCTGGCGTGGAGCACACCAAGAAATTCCTCCCTCAGTGGTTGCA 86
 Db 552 AATGTCCTTTAGGGCTCTGGAGAGAGCACACTGAAAGTTTGGCATCTTGGTTTATA 711
 QY 87 GCTGTCATCCGCTCTTCCATTCATTCGATGGAATCTGAGGAATCTGTCACATGCCCAAG 146
 Db 712 GCTCTTCACGACGCGTTCATTCATGAGGAATCTGAGGAATCTGTCACATGCCCAAG 771
 QY 147 ACTGCCATCGCATTCACCATAGCAGCTCCATTCGTTGGTTCAGACAAATCGGGTCGAGGGCG 206
 Db 772 ACAGCTATGGTATTCACCATAGCAGCATCAGTTTGGGACAGGTGATTGGGTCAAGAGCA 831
 QY 207 GAGCGATTCGTCGAGGCACTGGTGCACAGA 240
 Db 832 GAGAGCGGAGACTCAAGTCTGTAGCTGAGAGA 865

RESULT 8

AAC54760
 ID AAC54760 standard; DNA; 1346 BP.

XX AC AAC54760;

XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 78990.

XX KW Hybridisation assay; Genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway;
 KW metabolic pathway; promoter; termination sequence; ss.

XX OS Arabidopsis thaliana.

XX PN EPI033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 02-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.

Query Match	15.64	Score 105.2	DB 21	Length 1346
Best Local Similarity	68.24	Pred. No. 2e-22		
Matches 146	Conservative 0	Mismatches 68	Indels 0	Gaps 0
QY 27	AACTCCATCTTGGGGTGGTGGACACACCAAGAAATTCCTCCCTCAGTGGTTGCA	86		
DB 803	AATGTGCTTTAGGGTCTGGAGAGACCACTCAAAAGTTTCGGCATCTTGGTTTATA	862		
QY 87	GCTGTCCATGCCGTGTTCATTCATTTGGAATGCTGAGGAATCTGTCAACATGCCCCNAG	146		
DB 863	GCTCTTTCACGCGGTTCCTCATTCATAGGAATCTGAGAAAGTCAGTGTTCATGCTCTAAA	922		
QY 147	ACTGCCATGCATTCACCATAGCAGCCTCCATTTGTTGGTTCAGACAATCGGGTCGAGGCG	206		
DB 923	ACAGCTATGTTTACCATAGCAGCATCAGTTTGGACAGGTGATTGGTCAAGAGCA	982		
QY 207	GAGCGCATTCGTTGAGGCACTGGCTGCAAGA	240		
DB 983	GAGAGCGGAGACTCAAGTCTGTAGCTGAGAAGA	1016		
RESULT 9				
AAC35078				
ID	AAC35078 standard; DNA; 1349 BP.			
XX	AAC35078;			
AC	AAC35078;			
DT	17-OCT-2000 (first entry)			
XX	Arabidopsis thaliana DNA fragment SEQ ID NO: 8918.			
XX	Hybridization assay; Genetic mapping; gene expression control;			
KW	Protein identification; signal transduction pathway;			
KW	Metabolic pathway; promoter; termination sequence; ss.			
XX	Arabidopsis thaliana.			
OS	Arabidopsis thaliana.			
XX	EP1033405-A2.			
XX	06-SEP-2000.			
PD	25-FEB-2000; 2000EP-0301439.			
XX	25-FEB-1999; 99US-0121825.			
PR	05-MAR-1999; 99US-0123180.			
PR	09-MAR-1999; 99US-0123548.			
PR	23-MAR-1999; 99US-0125788.			
PR	25-MAR-1999; 99US-0126264.			
PR	29-MAR-1999; 99US-0126785.			
PR	01-APR-1999; 99US-0127462.			
PR	06-APR-1999; 99US-0128234.			
PR	08-APR-1999; 99US-0128714.			
PR	16-APR-1999; 99US-0129845.			
PR	19-APR-1999; 99US-0130077.			
PR	21-APR-1999; 99US-0130449.			
PR	23-APR-1999; 99US-0130510.			
PR	28-APR-1999; 99US-0130891.			
PR	30-APR-1999; 99US-0131449.			
PR	30-APR-1999; 99US-0132048.			
PR	30-APR-1999; 99US-0132407.			
PR	04-MAY-1999; 99US-0132484.			
PR	05-MAY-1999; 99US-0132485.			
PR	06-MAY-1999; 99US-0132486.			
PR	07-MAY-1999; 99US-0132487.			
PR	07-MAY-1999; 99US-0132863.			
PR	11-MAY-1999; 99US-0134256.			
PR	14-MAY-1999; 99US-0134218.			
PR	14-MAY-1999; 99US-0134219.			
PR	14-MAY-1999; 99US-0134221.			
PR	14-MAY-1999; 99US-0134370.			
PR	18-MAY-1999; 99US-0134768.			
PR	19-MAY-1999; 99US-0134941.			

PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139432.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 21-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140363.
PR 23-JUN-1999; 99US-0140364.
PR 24-JUN-1999; 99US-0140699.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149375.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 01-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158212.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.

```
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 15.6%; Score 105.2; DB 21; Length 1349;
Best Local Similarity 68.2%; Pred. No. 2e-22;
Matches 146; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 27 AAATCCATCTTGGGTGCGGTGAGCAGCAGCAGCAATCTCCCTCAGTGGTTTCCA 86
Db 803 AATGTGCTTTAGGGTCTGGAGAGCAGCAGTGAAGTTTTCGGCATCTTGTATTATA 862

QY 87 GCTGTCCATGCGGTGTTTCCATTTCATTTGGAATGCTGAGGAATCTGCAATGCCCAAG 146
Db 863 GCTCTTCAGCAGCGGTTCCATTTCATAGTAATGAGAAAGTCAGTTGATGCTCTAAA 922

QY 147 ACTGCCATGGATTACCATAGCAGCCTCCATTGTTGTCAGACATCGGTTCAGGGCG 206
Db 923 ACACGTATGTTATTCACCATAGCAGCATCAGTTTGGGACAGGTGATGGTCAAGAGCA 982

QY 207 GAGCGCATCTGCTGAAGGCACTGGCTGCAAGA 240
Db 983 GAGAGCGGAGACTCAAGTCTGTAGCTGAGAAGA 1016

RESULT 10
AAC37164
ID AAC37164 standard; DNA; 490 BP.
XX
AC AAC37164;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 16408.
XX
KW Hybridisation assay; Genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 23-JUL-1999; 99US-0134218.
PR 23-JUL-1999; 99US-0145224.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139760.
PR 18-JUN-1999; 99US-0139763.
PR 22-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142190.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
```

Query Match
Best Local Similarity 6.7%; Score 44.8; DB 21; Length 490;
Matches 61; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

QY 33 CATCTTGGGCGTGCGGTGAGCACCAAGAATTTCCTCCCTCAGTGGTTGCAGCTGTC 92
DB 403 CCTTTAGGGGCTCGAGAGACACTGAAAAGTTTTGCGCATCTTGGTTATAGCTCTT 462

QY 93 CATGCCGCTGTTCCATTCAITGGGAATGC 120
DB 463 CACGACGCGGTTCATTCTAGGGATAC 490

RESULT 11
ABL34112
ID ABL34112 standard; DNA; 7624 BP.
XX ABL34112;
AC ABL34112;
DT 26-MAR-2002 (first entry)
XX Human immune system associated gene SEQ ID NO: 2085.
DE Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
Gene, ds.
XX Homo sapiens.
OS WO200200928-A2.
PN 01-JAN-2002.
PD 02-JUL-2001; 2001WO-EP07537.
PF 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043626.
XX (EPIG-) EPIGENOMICS AG.
PA Olek A, Piepenbrock C, Berlin K;
XX WPI, 2002-130909/17.
PT Nucleic acid comprising fragment of chemically modified gene, useful
for diagnosis and treatment of diseases associated with abnormal
cytosine methylation -
XX Claim 1; SEQ ID NO 2085; 32pp + Sequence Listing; German.
XX The present invention provides a number of human immune system associated
genes which are modified by the methylation of cytosines. The sequences
can be used in the diagnosis and treatment of immune system disorders,
including eye diseases such as retinopathy, neovascular glaucoma and
macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
rheumatoid arthritis, psoriasis and inflammatory/neurotic bowel
disease.

CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
XX

SQ Sequence 7449 BP; 2244 A; 120 C; 1640 G; 3445 T; 0 other;
Query Match 5.9%; Score 39.6; DB 24; Length 7449;
Best Local Similarity 50.0%; Pred. No. 0.3;
Matches 99; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 469 TTAATCTGTGATGAATGCTTTCCTGTCCTCCAGAGTATTCGTTGTTGGGAT 528
Db 1162 TTTTAGTTTATGTAATTTTATTTTATTTTATTTTAAAGCAATTCGCTGTTAGTTTTT 1221
QY 529 GAGACAAGTGGTAAGAGTGCTACTATATACAGATCATTCGTTGTTAAAGTTGCCAG 588
Db 1222 AAGTAGTGGGATTATAGTGTGAGTTATTATTTAGTTAATTTTGTATTTTGGTAG 1281
QY 589 TTCTGCAGTTCATGTCATCTGTAATTTGATGATGCTGGATTTCTACTATTATCAATCCTC 648
Db 1282 AGATGGGATTTTGTATGTTAGTTAGTGGTTGGTTTGAATTTTATTTTAAAGCGATTTT 1341
QY 649 ATTACTGTGTGTAATA 666
Db 1342 TTTTTCGGTTTATAA 1359

RESULT 14
ABL70527
ID ABL70527 standard; DNA; 19459 BP.
XX AC ABL70527;
XX DT 01-JUL-2002 (first entry)
XX DE Chemically treated cell signalling DNA sequence#209.
XX DE Cell signalling; cytosine methylation; cell signalling disease;
KW cancer; tumour; cytostatic; ds.
XX OS Unidentified.
XX OS WO200202807-A2.
XX PN 10-JAN-2002.
XX PD 29-JUN-2001; 2001WO-EP07471.
XX PF 30-JUN-2000; 2000DE-1032529.
XX PR 01-SEP-2000; 2000DE-1043826.
XX XX (EPIG-) EPIGENOMICS AG.
XX XX Olek A, Piepenbrock C, Berlin K;
XX XX WPI; 2002-154758/20.
XX DR Nucleic acid, useful for diagnosis and therapy of diseases associated
XX PT with cell signalling e.g. cancer, comprises chemically modified genomic
XX PT sequences of genes associated with cell signalling.

XX Claim 1; SEQ ID NO 417; 24pp+sequence listing; English.
XX PS The invention relates to a nucleic acid comprising a sequence of at least
XX CC 18 bases of a segment of chemically pretreated DNA of genes associated
XX CC with cell signalling. The activity of the modified sequences of the
XX CC invention may be described as cytostatic. The object of the invention is
XX CC to provide the chemically modified DNA of genes associated with cell
XX CC signalling, as well as oligonucleotides and/or PNA-oligomers for

CC detecting cytosine methylations, as well as a method which is
CC particularly suitable for the diagnosis and/or therapy of genetic and
CC epigenetic parameters of genes associated with cell signalling. The
CC chemically modified DNA provided by the invention is useful for diagnosis
CC and therapy of diseases such as solid tumours and cancer. The sequences
CC given in records ABL70111-ABL70626 represent chemically pre-treated
CC genomic DNA's of genes associated with cell signalling.
CC Note: The sequence data for this patent is not represented in the printed
CC specification, but is based on sequence information supplied by the
CC European Patent Office.

SQ Sequence 19459 BP; 5585 A; 245 C; 3944 G; 9685 T; 0 other;
Query Match 5.9%; Score 39.4; DB 24; Length 19459;
Best Local Similarity 48.8%; Pred. No. 0.6;
Matches 106; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 438 TGGAGAGTGGAGAGTGATATTAAGTCTATTTATCTTGTGATGTAATTCCTTCCTTCGTTG 497
Db 19239 TGGATGAAGCTTGTAGTTATAGAAATTTGTTTTTAAAGATTGAAGTTGTTATTTTA 19298
QY 498 TCCCTCAGAGTATTCGTTTGTGTTGGGATGAGACAAGTGGATAGAGTGTCTACTATA 557
Db 19299 TTTGTTTAAAGTATAATTTAAATTAATTTGTTAGTATTTGTTTAAAGTATTTTGA 19358
QY 558 TACAGATCATCTCTGTTGTTAAAGTTTCCAGTTCTGCAGTTTCATGTAATTTGAT 617
Db 19359 ATAAAGCAGTTTATATGAGAAATTTATTTTATATATATTTTATTTATTTTAT 19418
QY 618 GATGCTGGATTTCTACTATTATCAATCGTCATTATA 654
Db 19419 AATTTGTTTATTTTATTTTAAATTTTAAATTTTAAATTTTA 19455

RESULT 15
ABK31212
ID ABK31212 standard; DNA; 19459 BP.
XX AC ABK31212;
XX DT 23-APR-2002 (first entry)
XX DE Signal transduction associated gene modified DNA #28.
XX DE Human; signal transduction associated gene; cytosine methylation state;
KW CpG island; signal transduction associated disease; solid tumour; cancer;
KW antitumour; cytostatic; mutant; ds.
XX OS Homo sapiens.
XX OS Synthetic.
XX OS WO200200926-A2.
XX PN 03-JAN-2002.
XX PD 29-JUN-2001; 2001WO-EP07472.
XX PF 30-JUN-2000; 2000DE-1032529.
XX PR 01-SEP-2000; 2000DE-1043826.
XX XX (EPIG-) EPIGENOMICS AG.
XX XX Olek A, Piepenbrock C, Berlin K;
XX XX WPI; 2002-147896/19.

XX PT Oligonucleotide for diagnosis and therapy of diseases associated with
XX PT signal transduction e.g. cancer, comprises chemically modified genomic
XX PT sequences of genes associated with signal transduction.
XX PS Claim 1; SEQ ID No 55; 24pp; English.
XX PS The present invention relates to chemically modified DNA sequences of
CC

CC signal transduction associated genes. The DNA sequences are chemically
 CC modified using a solution of bisulphite, hydrogen sulphite or
 CC disulphite. Also disclosed are oligonucleotides and/or RNA oligomers
 CC for detecting the cytosine methylation state (CpG islands) of these
 CC genes, and a method for the diagnosis and/or therapy of genetic and
 CC epigenetic parameters of genes associated with signal transduction.
 CC The genomic DNA can be obtained from cells or cellular components which
 CC contain DNA, e.g. cell lines, biopsies, blood, sputum, stool, urine,
 CC cerebral spinal fluid, tissue embedded in paraffin such as tissue from
 CC eyes, intestine, kidney, brain, heart, prostate, lung, breast or liver,
 CC histologic object slides, and all their possible combinations. The
 CC sequences of the invention are useful for the diagnosis and therapy of
 CC diseases associated with signal transduction e.g. solid tumours and
 CC cancer. ABX31159-ABX31545 represent chemically pretreated genomic DNA
 CC sequences of different genes associated with signal transduction, or
 CC their complementary sequences.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 XX
 SQ Sequence 19459 BP; 5585 A; 245 C; 3944 G; 9685 T; 0 other;

Query Match 5.9%; Score 39.4; DB 24; Length 19459;
 Best Local Similarity 48.8%; Pred No. 0.6;
 Matches 106; Conservative 0; Mismatches 111; Indels 0; Gaps 0;
 QY 438 TGGAGAGTGGAGAGTGTGATATGAGTCTATTTTCTTGTGATGTAATGCGCTTGCCTTG 497
 DB 19239 TGGATGAAGGTTTGTAGTATAGAAATTTGTTTTTATAGATTTCAAGCTTTGTATTTTA 19298
 QY 438 TCCCTCAGAGTATTCGTTGTTGTTGGGATGAGACAGTGAATAAGAGTGCTACTATA 557
 DB 19299 TTTGTTTAAAGTATTAATTTAAATAATTTTGTAGTATTTGTTTAAAGTATTTAGA 19358
 QY 558 TACACGATCATCTCTGTTGTTAAGTTTGGCAGTTCGCAAGTTCATGTCATCTGTAATTGAT 617
 DB 19359 ATAAAGGAGTTATTATGGAGAAATTTATTTTATATATATATATATATATTTTAT 19418
 QY 618 GATGCTGGATTTCTACTATTTTCAATCGTCATTATA 654
 DB 19419 AATTTGTTTATTTTATTAATTTTAAATTTTAAATTTTA 19455

Search completed: November 14, 2003, 08:14:46
 Job time : 229 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: November 14, 2003, 07:42:12 ; Search time 1919 seconds
(without alignments)
14347.149 Million cell updates/sec
Title: US-09-868-025-1
Perfect score: 673
Sequence: 1 ttacattgctgctcgat.....actgtgtgtaaaaaaaaaa 673

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 2888711 seqs, 2045481386 residues
Total number of hits satisfying chosen parameters: 5777422
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_hgt.*
3: gb_in.*
4: gb_on.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_on.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vl.*
30: em_hgt_hum.*
31: em_hgt_inv.*
32: em_hgt_other.*
33: em_hgt_mus.*
34: em_hgt_pln.*
35: em_hgt_rtd.*
36: em_hgt_mam.*
37: em_hgt_vrt.*
38: em_sy.*
39: em_hgtgo_hum.*
40: em_hgtgo_mus.*
41: em_hgtgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	673	100.0	673	8	AF192975	AF192975 Oryza sat
2	630.6	93.7	123673	8	AP002837	AP002837 Oryza sat
3	467.6	69.5	80327	2	AP003955	AP003955 Oryza sat
4	467.6	69.5	191580	8	AP003847	AP003847 Oryza sat
5	323.4	48.1	1184	6	AX653397	AX653397 Sequence
6	121	18.0	1060	8	BT002755	BT002755 Arabidops
7	121	18.0	1180	8	AY050908	AY050908 Arabidops
8	121	18.0	1275	8	AY088538	AY088538 Arabidops
9	121	18.0	85020	6	AB018113	AB018113 Arabidops
10	120.6	17.9	1104	6	AX654660	AX654660 Sequence
11	120.6	17.9	95893	8	AE017123	AE017123 Oryza sat
12	120.6	17.9	139999	8	AC018727	AC018727 Oryza sat
13	116.4	17.3	148522	2	AC126015	AC126015 Medicago
14	105.2	15.6	1035	6	AX507557	AX507557 Sequence
15	105.2	15.6	1035	6	AX651340	AX651340 Sequence
16	105.2	15.6	1349	8	AY085639	AY085639 Arabidops
17	105.2	15.6	99688	8	ATP13M23	AL035523 Arabidops
18	105.2	15.6	197070	8	ATCHRIV62	AL161562 Arabidops
19	60.4	9.0	2000	6	AX655393	AX655393 Sequence
20	43.8	6.5	7624	6	AX347014	AX347014 Sequence
21	42	6.2	7634	6	AX347032	AX347032 Sequence
22	41.8	6.2	162525	2	AC140026	AC140026 Medicago
23	41.6	6.2	181067	2	BX005309	BX005309 Danio rer
24	41.2	6.1	349980	6	AX344560	AX344560 Sequence
25	41	6.1	176577	5	AL929171	AL929171 Zebrafish
26	41	6.1	187998	2	AC118680	AC118680 Mus muscu
27	41	6.1	300087	2	AL528900	AL528900 Danio rer
28	40.6	6.0	40031	3	CBRG01D9	U56488 Caenorhabdi
29	40.4	6.0	125020	9	AF429315	AF429315 Homo sapi
30	40.4	6.0	127917	8	AC118673	AC118673 Genomic s
31	40.4	6.0	145146	10	AL929187	AL929187 Mouse DNA
32	40.4	6.0	156933	8	AC125411	AC125411 Genomic s
33	40.4	6.0	162241	8	AF485811	AF485811 Oryza sat
34	40.4	6.0	189349	10	AC117190	AC117190 Mus muscu
35	39.8	5.9	164949	2	AC119839	AC119839 Mus muscu
36	39.8	5.9	204113	2	AC126670	AC126670 Mus muscu
37	39.6	5.9	7449	6	AX345178	AX345178 Sequence
38	39.6	5.9	193154	5	AL929006	AL929006 Zebrafish
39	39.4	5.9	19459	6	AX344208	AX344208 Sequence
40	39.4	5.9	19459	6	AX348959	AX348959 Sequence
41	39.4	5.9	108100	9	AC103881	AC103881 Homo sapi
42	39.4	5.9	128809	2	AC121494	AC121494 Mus muscu
43	39.4	5.9	141262	9	AC080033	AC080033 Homo sapi
44	39.4	5.9	169059	9	AC009514	AC009514 Homo sapi
45	39.4	5.9	169060	9	AC087856	AC087856 Homo sapi

ALIGNMENTS

RESULT 1
AF192975
LOCUS AF192975 673 bp DNA linear PLN 09-NOV-1999
DEFINITION Oryza sativa unknown gene.
ACCESSION AF192975
VERSION AF192975.1 GI:6289051
KEYWORDS Oryza sativa (indica cultivar-group)
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoidae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 673)
AUTHORS Patel,V.M., Mathai,C.A., Divya,C. and Ashok,M.
TITLE Oryza sativa Variety IR64 (CDNA clone AGTSAU-11 from 7 days old

seedling)
Unpublished
2 (bases 1 to 673)
Patell V.M., Mathai, C.A., Divya, C. and Ashok, M.
Direct Submission
Submitted (08-OCT-1999) Plant Genome Biology Department, Avesthagen
Graine Technologies, P.Box 5091, Cubbon Park G.P.O., Bangalore,
Karnataka 560001, India
Location/Qualifiers
1..673
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="IR64"
/db_xref="taxon:39946"
/clone="AGIRAL-11"
/tissue_type="7 day old seedling"
complement(1..431)
/product="unknown"
complement(1..431)
/function="role in salinity stress response"
/codon_start=1
/product="unknown"
/protein_id="AF06789.1"
/db_xref="GI:6289052"
/translation="MVDTFPISEQSETHAMWSSSTTRPSRHLHRRIPCLALGV
TAICSLVWIVHSVGGGIVLSCQCLQTNALRPDCLTNNGCYGEGHSLGHVDRF
PQHSNWSGMDCKPLRGFLGVLTPHPEFAAIRAGKV"
BASE COUNT 163 a 145 c 170 g 195 t
ORIGIN
Query Match 100.0%; Score 673; DB 8; Length 673;
Best Local Similarity 100.0%; Pred. No. 1.1e-182;
Matches 673; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTTACCTTGCTGCTCGATGGCAGCAACTCCATCTTGGGGTGTGGCGTGAGCACACCA 60
DB 1 TTTACCTTGCTGCTCGATGGCAGCAACTCCATCTTGGGGTGTGGCGTGAGCACACCA 60
QY 61 AGAAATTCCTCCCTCAGTGTTCGACGTGTCATGCGCGTGTTCATTCATTTGGATGC 120
DB 61 AGAAATTCCTCCCTCAGTGTTCGACGTGTCATGCGCGTGTTCATTCATTTGGATGC 120
QY 121 TGAGGAATCTGCAACATGCCAAGACTGCATGGCAATTCACCATAGAGCCTCCATTC 180
DB 121 TGAGGAATCTGCAACATGCCAAGACTGCATGGCAATTCACCATAGAGCCTCCATTC 180
QY 181 TTGTCAGACAATCGGTTCGAGGCGGAGCGCATTCGCTGAAGCACTGGCTGCAAGA 240
DB 181 TTGTCAGACAATCGGTTCGAGGCGGAGCGCATTCGCTGAAGCACTGGCTGCAAGA 240
QY 241 GCGAGCTGATTCACCAACCGCTGACATGTATCCAAACAGACTGCAAAATTCGAGTG 300
DB 241 GCGAGCTGATTCACCAACCGCTGACATGTATCCAAACAGACTGCAAAATTCGAGTG 300
QY 301 ACACGAGGGAAGGATGGATCGCTCGCGATGAAGATGATGGCGGACGGGCTTCG 360
DB 301 ACACGAGGGAAGGATGGATCGCTCGCGATGAAGATGATGGCGGACGGGCTTCG 360
QY 361 GTGGTGTGCTGCTCCACCAACAGCATGTGTTCTGATTCCTCAGTGGAAATTT 420
DB 361 GTGGTGTGCTGCTCCACCAACAGCATGTGTTCTGATTCCTCAGTGGAAATTT 420
QY 421 GTATCTACCATATCCCTCGAGAGTGGAGTGTATATGATGTATTTATCTTGTGAT 480
DB 421 GTATCTACCATATCCCTCGAGAGTGGAGTGTATATGATGTATTTATCTTGTGAT 480
QY 481 GTAATTCCTTTGCTGCTCCCTCAGAGTATTCGTTGTTGTGGATGACAGAGTGA 540
DB 481 GTAATTCCTTTGCTGCTCCCTCAGAGTATTCGTTGTTGTGGATGACAGAGTGA 540
QY 541 ATAAGAGTCTACTATATACAGCATCTCTGTTGTTAAAGTTTGCAGTTCTGCAATCA 600
DB 541 ATAAGAGTCTACTATATACAGCATCTCTGTTGTTAAAGTTTGCAGTTCTGCAATCA 600

QY 601 TGATCTGTAATTTGATGCTGGATTCTTACTATTATCAATCGTCATTACTGTGT 660
DB 601 TGATCTGTAATTTGATGCTGGATTCTTACTATTATCAATCGTCATTACTGTGT 660
QY 661 GTAAATAAAAAA 673
DB 661 GTAAATAAAAAA 673
RESULT 2
AP002837/c 123673 bp DNA linear PLN 26-OCT-2002
LOCUS Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 6,
DEFINITION BAC clone:OSJNBa0019F11.
ACCESSION AP002837
VERSION AP002837.2 GI:24413940
KEYWORDS Oryza sativa (japonica cultivar-group)
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehretidae; Oryzaceae; Oryza.
1
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, BAC
clone:OSJNBa0019F11
Published Only in Database (2000)
2 (bases 1 to 123673)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Direct Submission
Submitted (03-AUG-2000) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki, 305-8602, Japan
(E-mail:tsukuba.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)
On Oct 26, 2002 this sequence version replaced gi.9711842.
Genes were predicted from the integrated results of the following:
GENSCAN (http://ccr-081.mit.edu/GENSCAN.html), FGENESH
(http://www.softberry.com/), GeneMark.hmm
(http://opal.biology.gatech.edu/GeneMark/), GlimmerM
(http://opal.biology.gatech.edu/glimmer/glmr_form.html), RiceHMM
(http://rgp.dna.affrc.go.jp/RiceHMM/), SplicePredictor
(http://bioinformatics.iastate.edu/cgi-bin/sp.cgi), BLASTN and
BLASTX. The genomic sequence was searched against NCBI Nonredundant
Protein database, nr
(ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at
RGP. Protein homologies of the coding regions were searched against
NCBI Nonredundant Protein database with BLASTP. ESTs represent the
identified cDNA sequences using BLASTN with the corresponding DBJ
accession no. and RGP clone ID.
A gene with identity or significant homology to a protein is
classified based on the protein name to indicate the homology level
such as same name, 'putative', and '-like protein'. A gene without
significant homology to any protein but with EST homology (covering
almost the entire length of partial sequence) is classified as an
'unknown' protein. A gene predicted by two or more gene prediction
programs is classified as a 'hypothetical' protein according to
IRGSP standard. A gene predicted by a single gene prediction
program is also classified as a probable 'hypothetical' protein and
is included as a miscellaneous feature of the sequence.
The orientation of the sequence is from -21M13 to M13rev of the
BAC clone. This sequence of OSJNBa0019F11 clone has an overlap with
P0541H01 (DBJ:AP001389) clone at the 5' end. Detailed information
on overlap and assembly quality together with annotation of this
entry is available at
http://rgp.dna.affrc.go.jp/Genomeseq.html.
FEATURES
Location/Qualifiers
1..123673
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"

Db 25585 CACCCCGTGGCTGACATGTATCCAAACAAGTCCTGGCAATTGCAGTGACACTGAGGCCAA 25522

Qy 314 GCGATGGGATCCGCTCGGATGAGATGATCGCGGAGCGGCTTCTGGTGGTGGTGGTGC 373

Db 25525 GGCATGGGATCCGCTCGGAGTAA---GATGGCTGGACGGGCTTCTGGTGGTGGTGGTGC 25469

Qy 374 TCCAACACCACCAAGCATGTGTTTCTGATTGCTCAGTGATTGGAAAA---TTTGATATCTACCA 430

Db 25468 ACCAACACCACCAAGCATGTGTTTCTGATTGCTCAGTGATTGGAAAAATTGTGCTCATACCA 25409

Qy 431 GTATCCCTGGAGAGTGGAGATGGATATTTAGTGCTA-TTTTATCTTGTGATGTAAATTGCC 489

Db 25408 GTAATCTGTATGTTTGGAGATGATATTTAGTCTATTTTTATCTGTGATGTAAATTGCC 25349

Qy 490 TTTCGTTGTCCTCAGAAGTATTCCTTTGTTTGGGATGACAGCAAGTGGGAATTAAGATG 549

Db 25348 TTTCCTTTGCCCTCAGAAGTATTCCTTTGTTTGGGATGACAGCAAGTGGGAATTAAGATG 25289

Qy 550 CTACTATATACAGATCATCTCTGTTTAAAGTTTCCAGTTCTCGAGTTCAATGTATCTGT 609

Db 25288 CTGC-----AGTTCTGCAAGTTAAGCTTGCCAGTTCTCGAGTTCAATGTATCTGT 25241

Qy 610 AATTGATGATGCGAATTTCTACTA-TTTATCAATCTCATATTACTTCTGTATAA 664

Db 25240 AATTGATGTCGAGATTTCTGCTATTATTACCAATCATCATTTGACTGTGTATAA 25185

RESULT 5.

AX653397 LOCUS AX653397 1164 bp DNA linear PAT 22-MAR-2003

DEFINITION Sequence 3267 from Patent WO03000898.

ACCESSION AX653397

VERSION AX653397.1 GI:29156211

KEYWORDS

SOURCE OrYZa sativa

ORGANISM OrYZa sativa

REFERENCE 1 Chang,H.S.; Chen,W.; Cooper,B.; Glazebrook,J.; Goff,S.A.; Hou,Y.M.; Katagiri,F.; Quan,S.; Tao,Y.; Whitham,S.; Xie,Z.; Zhu,T. and Zou,G.G. Plant genes involved in defense against pathogens Patent: WO 03000898-A 3267 03-JAN-2003;

JOURNAL Syngenta Participations AG (CH)

FEATURES

Source Location/Qualifiers

1..1164 /organism="OrYZa sativa"

/mol_type="genomic DNA"

/db_xref="taxon:4530"

BASE COUNT 294 a 286 c 297 g 287 t

ORIGIN

Query Match 48.1%; Score 323.4; DB 6; Length 1164;

Best Local Similarity 93.5%; Pred. No. 7,2e-82;

Matches 360; Conservative 0; Mismatches 21; Indels 4; Gaps 2;

Qy 15 TCAGATGGCAGCAAAC-TCGATCTGGGTTGGGTGAGCACACGAGAATTTCTCCC 73

Db 783 TGGAAATGGCAGCAAAATCTCTCGGGGTGGGCGTAGCACACGAGAATTTCTCCC 842

Qy 74 TCAGTGGTTTGCAGCTGTCCATGCCGCTGTTCCATTCATTGGAATGCTGAGGAATCTGT 133

Db 843 GCAGTGGTTTGCAGCTGTCCATGCCGCTGTTCCATTCATTGGAATGCTGAGGAATCTGT 902

Qy 134 CAACATGCCCAAGACTGCCATGSCATTCACCATAGCGCTCCATTTGTTGTCAGCAAT 193

Db 903 GAACATGCCCAAGACTGCCATGSCATTCACCATAGCAGCTCCATTTATTTGGCAGCAAT 962

Qy 194 CGGTCGAGGCGGAGCGCAATTCGTCGTAGGCGACATGGCTGCAAAAGAGCGAGCTGATTC 253

Db 963 CGGTCGAGGCGGAGCGCAATTCGTCGTAGGCGACATTCGTCGTGCAAAAGGCGAGCGGATTC 1022

QY 254 CACCACCGTGGTGCATCTATCCAAAGACTGCARATTCAGTGCACACGAGGGCAA 313
 Db 1023 CACCACCGTGGTGCATCTATCCAAAGACTGCARATTCAGTGCACACGAGGGCAA 1082
 QY 314 GGCATGGATCCGCTCCGATGAAGATGATGCGGACGGCTTCGTGGTGGTCTGCTGC 373
 Db 1083 GGCATGGATCCGCTCCGATGAAGATGATGCGGACGGCTTCGTGGTGGTCTGCTGC 1139
 QY 374 TCCACACCAACGATGTTCTTGA 398
 Db 1140 ACCAACACCAAGCATGTGTTCTGA 1164

RESULT 6
 BT002755 1060 bp mRNA linear PLN 15-JAN-2003
 LOCUS Arabidopsis thaliana clone C104940 unknown protein (At5g45410)
 DEFINITION mRNA, complete cds.
 ACCESSION BT002755
 VERSION 1
 KEYWORDS FLI CDNA.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 1 (bases 1 to 1060)
 Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,
 Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C.,
 Yu,G., Yuan,S., Chen,H., Cheuk,R., Jones,T., Kim,C.J., Nguyen,M.,
 Palm,C.J., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Davis,R.W.,
 Ecker,J.R., and Theologis,A.
 Arabidopsis Open Reading Frame (ORF) Clones
 Arabidopsis Unpublished
 2 (bases 1 to 1060)
 Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,
 Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C.,
 Yu,G., Yuan,S., Chen,H., Cheuk,R., Jones,T., Kim,C.J., Nguyen,M.,
 Palm,C.J., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Davis,R.W.,
 Ecker,J.R., and Theologis,A.
 Direct Submission
 Submitted (15-JAN-2003) Plant Gene Expression Center, 800 Buchanan
 Street, Albany, CA 94710, USA
 COMMENT Annotation based on July 2002 version of the Arabidopsis genome
 submitted to Genbank.

FEATURES
 source
 1..1060
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /chromosome="5"
 /clone="C104940"
 /note="This clone is in pUNI 51."
 ecotype="Columbia"
 gene
 1..1060
 /gene="At5g45410"
 CDS
 1..1029
 /gene="At5g45410"
 /codon_start=1
 /evidence=experimental
 /product="unknown protein"
 /protein_id="AAC22584.1"
 /db_xref="GI:27754261"
 /translation="MOECPPFAAESIIKCPFLRNINKPTNLSPSLFPPIPVQGGK
 PIFDGFDFDSAPKLFHKGIVPLSGPADDEAGRALQFNPLAGKVAITLSAF
 PGCGPFGPFGSKKKQKKKPKSNQSGDSKHEAVGDEWLNKNGCNPIAKSPRAA
 SKVPLISKALTLPQGMKYRCAPAVARAALSKTALVKSLRPOPLPKMLAIALMGV
 AANVPLVGRHTKPKSPAWFLAVHAAPFIAMLRKSLMPKTMALITGASILGQVI
 GSAERYRLKAVAKMVPVTAMVSGVNSQSPGSGISGGHCGEGGVQVYNNVIGES
 AKSTGLUCY"
 misc_difference 375
 /gene="At5g45410"

/note="compared to genomic sequence"
 /replace="a"
 1030..1060
 /gene="At5g45410"
 278 a 224 c 255 g 303 t
 BASE COUNT
 ORIGIN

Query Match 18.0%; Score 121; DB 8; Length 1060;
 Best Local Similarity 68.9%; Pred. No. 1.6e-23;
 Matches 166; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 27 AACTCCATCTTGGGTGGTGGAGACACCAAGAAATTCCTCCCTCACTGGTTGCA 86
 Db 661 AATGTGCTCTAGTGTATGAGAGACACACAAAGAAGTTTCGCTCGCATGTTTG 720
 QY 87 GCTGTCAATGCGCTGTTCATTCATTCGAGGAATCTGTCAACATCCCAAG 146
 Db 721 GCTGTCAATGCTGTGCGCTTTTATAGCATGTTAGGAATCTGTCATGCCCAA 780
 QY 147 ACTGCCATGCAATTCACATAGCAGCTCCATTGTTGGTCAGACAATCGGGTCAGGGCG 206
 Db 781 ACAGCCATGGCTTGGACCATTTGGAGCTTCGATCTTGGGACAGGTGATTGGGTCAAGAGCT 840
 QY 207 GAGCGCATTCCTCTGAAGGCACTGGCTGCAAGAGCGACGCTGATCCACCACCGTGGCT 266
 Db 841 GAGGTTACCGTCTCAAGCAGTAGCTAGAGAAATGTTCTGTACCGCTATGTTTCT 900
 QY 267 G 267
 Db 901 G 901

RESULT 7
 AY050908 1180 bp mRNA linear PLN 18-SEP-2002
 LOCUS Arabidopsis thaliana unknown protein (At5g45410) mRNA, partial cds.
 DEFINITION Arabidopsids
 ACCESSION AY050908
 VERSION 1
 KEYWORDS FLI CDNA.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 1 (bases 1 to 1180)
 Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K.,
 Goldsmith,A.D., Lee,J.M., Quach,H.L., Toriumi,M., Yu,G., Bowser,L.,
 Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T.,
 Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B., Lin,J.,
 Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M.,
 Seki,M., Shinn,P., Southwick,A., Tracy,S.E., Shinozaki,K., Davis,R.W.,
 Ecker,J.R., and Theologis,A.
 Arabidopsis Full Length cDNA Clones
 Arabidopsis Unpublished
 2 (bases 1 to 1180)
 Yamada,K., Liu,S.X., Pham,P.K., Banh,J., Banno,F., Dale,J.M.,
 Goldsmith,A.D., Jiang,P.X., Lee,J.M., Onodera,C.S., Quach,H.L.,
 Tang,C.C., Toriumi,M., Yamamura,Y., Yu,G., Yu,S., Bowser,L.,
 Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T.,
 Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Koesema,B., Lam,B.,
 Lin,J., Meyers,M.C., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J.,
 Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tracy,S.E.,
 Shinozaki,K., Davis,R.W., Ecker,J.R., and Theologis,A.
 Direct Submission
 Submitted (24-JUL-2001) Plant Gene Expression Center, 800 Buchanan
 Street, Albany, CA 94710, USA
 COMMENT RIKEN Genomic Sciences Center (GSC) members carried out the
 collection and clustering of RAPL cDNAs (RAPL CDNA: 'RIKEN
 Arabidopsis Full-length cDNA'): Seki,M., Narusaka,M., Ishida,J.,
 Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J., Hayashizaki,Y.,
 and Shinozaki,K.

The Salk, Stanford, PGEC (SSP) Consortium members carried out the sequencing and annotation of the RAPL cDNAs: Yamada, K., Liu, S.X., Pham, P.K., Bann, J., Banno, P., Dale, J.M., Goldsmith, A.D., Jiang, P.X., Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Yamamura, Y., Yu, G., Yu, S., Bowser, L., Chen, H., Cheuk, R., Jones, T., Karlin-Neumann, G., Kim, C., Koesema, E., Lam, B., Lin, J., Meyers, M.C., Miranda, M., Nguyen, M., Palm, C.J., Shinn, P., Southwick, A., Tracy, S.E., Davis, R.W., Ecker, J.R. and Theologis, A.

Yamada, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGEC) contributed equally to this work as PIs.

Annotation is based on the January 2002 version of the Arabidopsis genome submitted to GenBank.

FEATURES

source
Location/Qualifiers
1..1180
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="5"
/clone="RAPL08-14-C01 (R11029)"
/note="this clone is in a modified pBluescript vector (FLC-1) as a BamHI/XhoI insert."
ecotype: Columbia
<1..1180
/gene="At5g45410"
<1..998
/gene="At5g45410"
/codon_start=3
/evidence="experimental"
/product="unknown protein"
/protein_id="AAK93585.2"
/db_xref="GI:21954084"
/translation="ITQKCPFLRNINKPTNLSPFSLPFPVQGGKPIFEDGGPDS AFKLPHGKGIIVLPSGADSEAGRAALQFNPLAGKATISLSAFGGPFGFPF SEKKVQKQKPKSPKQVQSGDSKHEAVGDEWLKTCNCPKIAKSFRAAKWPLISKAL TLPPGKWKPCAPVAAALSKTALVKSLRPOPLPERKMLALMGAAVPLGVYRE HYKPSAPFLAVRAVPTLMLKSLVLMKPTAMALTICASITLGVISRAERYLKA VAEKMPVPTAMVSGVYQSGDSIGSGHGIGEGVKQVYVYNNISAKSYGLCY"
misc_difference 1
/gene="At5g45410"
/note="compared to genomic sequence"
/replaces="g"
999..1180
/gene="At5g45410"
1164
misc_feature
/gene="At5g45410"
/note="artifact within poly A tail"
BASE COUNT 330 a 232 c 274 g 344 t
ORIGIN

Query Match 18.0% Score 121; DB 8; Length 1180;
Best Local Similarity 68.9%; Pred. No. 1.6e-23;
Matches 166; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
QY 27 AACTCCATCTGGGGTGGGTGAGCAGACACCAAGAAATTCCTCCCTAGTGGTTGCA 86
DB 630 AATGTGCTCTAGTGTATGAGAGACACACAAAGATTTTCGCTGCTGATGTTTGG 689
QY 87 GCTGTCCATCCGCTGTTCCATTCATGATGATGAGGAATCTGTACATGCCCAAG 146
DB 690 GCTGTCCATGCTGCTGCTCTTTATAGCAATGCTTAGGAATCTGTTCTGATGCCCAA 749
QY 147 ACTGCCATGGCATTCAACATAGCAGCCTCCATTGTTGGTCAGACAATCGGGTGCAGGGCG 206
DB 750 ACAGCCATGGCTTTGACCATTTGGAGCTTCGATCTTGGACAGAGTATTGGGTCAAGAGCT 809
QY 207 GAGGCGATTCGCTGAGGCACTGGCTGCAAGAGCGACCTGATTCACACCGTGGCT 266
DB 810 GAGGCTTACCGCTCTCAAGAGCAGTAGCTGAGAAATGTTCTCTGTAAACCGCTATGTTCT 869
QY 267 G 267

Db 870 G 870

RESULT 8

AY088538
LOCUS Arabidopsis thaliana clone 767 mRNA linear PLN 14-APR-2003
DEFINITION Arabidopsis thaliana (thale cress)
ACCESSION AY088538
VERSION AY088538.1 GI:21407312
KEYWORDS Full cDNA.
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
REFERENCE
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 1275)
Haas, B.J., Volkovsky, N., Town, C.D., Troupkhan, M., Alexandrov, N., Feldmann, K.A., Flavell, R.B., White, O. and Salzberg, S.L.
Full-length messenger RNA sequences greatly improve genome annotation
Genome Biol. 3 (6), RESEARCH0029 (2002)
20988475
1209376
2 (bases 1 to 1275)
Brover, V., Troupkhan, M., Alexandrov, N., Lu, Y.-P., Flavell, R. and Feldmann, K.
Full-length cDNA from Arabidopsis thaliana
Unpublished
3 (bases 1 to 1275)
Brover, V., Troupkhan, M., Alexandrov, N., Lu, Y.-P., Flavell, R. and Feldmann, K.
Direct Submission
Submitted (11-MAR-2002) Ceres, Inc, 3007 Malibu Canyon Road, Malibu, CA 90265, USA
This clone sequence is one of 5,000 Ceres full-length cDNAs made available to TIGR and GenBank. The following quality assessment of this set was done by comparison with known proteins: two percent of the clones are estimated to be 5'-truncated, less than one percent are 3'-truncated, approximately two percent represent alternative splice variants, including unspliced introns and spliced exons; one percent may contain premature stop codons; five percent may have frame shifts in a coding region. A sequence is considered to be 5'-truncated if it lacks the translation initiation start (ATG). A sequence is considered to be 3'-truncated if it lacks the C-terminal end of the encoded protein. Please note that these cDNA sequences are derived from the Ws or Laer ecotypes and therefore may contain polymorphisms when compared to sequences from Col-0. Genet carried out the library production and sequencing of the full-length clones. Ceres, Inc. carried out the clustering of the 5' sequences, selection of clones, and sequence assembly.
Location/Qualifiers
1..1275
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="767"
81..1112
/codon_start=1
/product="unknown"
/protein_id="AAK66071.1"
/db_xref="GI:21595083"
/translation="MDQECPEAESIIQKCPFLRNINKPTNLSPFSLPFPVQGGKGIIVLPSGADSEAGRAALQFNPLAGKATISLSAFGGPFGFPF SEKKVQKQKPKSPKQVQSGDSKHEAVGDEWLKTCNCPKIAKSFRAAKWPLISKAL TLPPGKWKPCAPVAAALSKTALVKSLRPOPLPERKMLALMGAAVPLGVYRE HYKPSAPFLAVRAVPTLMLKSLVLMKPTAMALTICASITLGVISRAERYLKA VAEKMPVPTAMVSGVYQSGDSIGSGHGIGEGVKQVYVYNNISAKSYGLCY"

FEATURES

source
Location/Qualifiers
1..1275
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="767"
81..1112
/codon_start=1
/product="unknown"
/protein_id="AAK66071.1"
/db_xref="GI:21595083"
/translation="MDQECPEAESIIQKCPFLRNINKPTNLSPFSLPFPVQGGKGIIVLPSGADSEAGRAALQFNPLAGKATISLSAFGGPFGFPF SEKKVQKQKPKSPKQVQSGDSKHEAVGDEWLKTCNCPKIAKSFRAAKWPLISKAL TLPPGKWKPCAPVAAALSKTALVKSLRPOPLPERKMLALMGAAVPLGVYRE HYKPSAPFLAVRAVPTLMLKSLVLMKPTAMALTICASITLGVISRAERYLKA VAEKMPVPTAMVSGVYQSGDSIGSGHGIGEGVKQVYVYNNISAKSYGLCY"

CDS

BASE COUNT 350 a 256 c 296 g 373 t
ORIGIN

Query Match

18.0%; Score 121; DB 8; Length 1275;

source
CDS
CDS
CDS
CDS
CDS

[illegible]

```

ELEKTKKNDDEG"
join(15757. 16151,16482. 16992)

```


/evidence=not experimental
/product="ATP-dependent Clp protease-like protein"
/protein_id="BAB09167.1"
/db_xref="GI:9758729"
/translation="MGLTSLSSSLKPSLVSRSLNLSASSASSSPKNNLYLPKLI
SPPLTTPSPPLRANASISMSQCSAIRGAESDWMGLLRERIVFLGSSIDFVAD
AIMSOLLIDAKPKDKIKFINSFGSLRLNVIWVMSATMAIYVQVLRVADSTI
ALGIAASASIIILGAGTKRPAEMTRMHIQPLGASGQAIIDVEIQAEVWNNKN
VTSIAGTSRSFEOLKIDRDRYMSPIEAIVEYGLIDGVIGDSIIPLEPVPDRKP
RVNVEISKDPMKRTPLPDEIV"
join(18396..18688,18782..21050)
/note="gene_id:MFC19.7"
/codon_start=1
/evidence=not experimental
/product="replication protein A1-like"
/protein_id="BAB09168.1"
/db_xref="GI:9758730"
/translation="NAVSLTEGVVMKMLNGEVTSETDMKPVLOVTEKLQSLKHQ
ESSNKYKLLSDGTLAGMLTSLNLSVNGTILQSGVIRLTAVICNLITRIIVI
MOLEIVKCNIIIGAPFSSINIFQSGVNTQNGSSEQQARSVDNGRGVESA
NSPQVHNSSDAGRYCVANSPOVWSSSDAGRYGVSANSFORQVWNRNSDAG
YQPOVSRVYSGSPSTPRVYSSNAGYGGSDQDPRAPTATTAISRPPVQSAI
QPOQPMVNRGPAVARNAPRINPIAALNPYQGRWTIKRVTSKADLRFFNPRGEG
KLFSDLLDADGERTVTCFNDVQDFDKIVGVNVLISRLNLPKQNFNHLPNDY
EHLDSASTIQCEDDGTIPRVYHFHFNIGDIENMNSTTDVIGIVSISPTVAIME
KNLTVQKESLQKXMSGESVEVTWGMFCNAEGOKLONLSDSGVPPVLAALKRIGE
AOKDEKLTSGKPMWTCATISFMKVENFCTACPLMGDRPSCSKVTVNNGDGTWR
CEKCKVDECDYRILQIQLODHLDTWATFOBAGEBIEGMMSAKDILYKYNQDE
EFEDIRVATFKIFKIKIETYSDEQRYKATVKAELNYSNTRFMLEIDKL
KIGDANSILKALTIKPSNNSVDADFNDSVGTGRTASVDARFGLPAANOVQYGNQY
SDARSGLGFTSCNCRSHSVNACPTLMSBPQGYMGCTNAGCGMFRQHVGSY"
complement(join(21477..22112,22203..22481,22575..22688))
/note="gene_id:MFC19.8"
pir||T05524
similar to unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="BAB09169.1"
/db_xref="GI:9758731"
/translation="MDQECPPFAAESIQKCPFLRNKINLPFLSPSSLGPPIPVQGGK
PIFEDGPGDSAFKFLPHGKGLVLPSSGADDEAGRALQFNPLAGKATISLAF
GGGPGPGFSEKWKQKQKPKPKNQSGSSSHVAGVDEWLTGCTPAKSRRA
SKVPLIISKALTLPQMKRYCPAPIVAAKALSKALVSLRPQDJPBKNLIALMGH
AANVPLGVNRETHKPSFAPLVAHAAVPIFAMLRKSLVLPKTMALITGASILQV
GSRABRYRLKAAEKVPTAMVSYNQSPGDSGSGHCGIGEGVKQVYNNVNGES
AKSTGLCY"
complement(25074..26132)
/note="gene_id:MFC19.9"
unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="BAB09170.1"
/db_xref="GI:9758732"
/translation="MGFTIPXNLRITWFKSPKTLISEIFPSKFLYLLQLLQCNRA
DPDEDPRPFRFOSRPSSTHAEERAPKPLFTISIGVSLIILSLISFFYFSESE
AKSLILSLFVAGPAPSSLTGKIRVGOILEPEOIHDESSDNRERESRKSVN
KRKSGTSDNPNENASAVTEVRKVIPOSKESGVNETHKMTABEILKOLIKH
PAGKRGWETVAFAGRYKTERVVKAEIIEGKLYESDSDTAQFLKRNKASDPLVD
ENENSCAGDASGTEIENWSGEDIALNLAKAPPEAAWRWEKIAAAVPGKSKAACM
KRYTELKXGFRSSKTPAN"
join(28606..28613,28709..28787,28846..28938,29022..29111,
29195..29333,29480..29582,29745..29797,29889..29955,
30054..30162,30258..30828,30906..31038,31082..31229)
/note="gene_id:MFC19.10"
/codon_start=1
/evidence=not experimental
/product="serine/threonine-protein kinase Max (male germ
cell-associated kinase)-like protein"
/protein_id="BAB09171.1"
/db_xref="GI:9758733"
/translation="MEKYLKKEVGDGTGNNVARNVKNOTNEVGTVEFVVOVALKEMK
KKYFSEECVNLREKSLSRNHPNVLKREIVRENDILYFVEFYMECNLYQLMDRPP

KHPAESDIRMWCQVFEQLSYMHQRYFHRDLKPNLNVKOVIKIADLGLAREIDSS
PPYTVSTRWYRAPEVLQSYVYTVSKVDMWAMGAIMBELLRLRPLFPGASEADELYK
ICSVIGSTPEETWLEGLNLASVINYQPPQPGVHLSVMPYASADAVNLIERLCSWDP
CNRPTABALQHPFQSCYIIPSLRPLSEPRGSLQSQVKRLPAAFTYNAKPLN
TVYTPNTAPFSGVNSVYTPKTNAPFSGVNSVYTAKTAPFSGVNSVYNAKPLN
PDNGATQCKVDVNTNNKONTANNKPVRSVNVKDSKYRPPGRKSPLAASMEKNWFSR
GASETANKLANATIGRWTSQRQOQPAKAGLPIPTLGSVPYKLDVMAHLSYBES

Query Match 18.0%; Score 121; DB 8; Length 85020;
Best Local Similarity 68.9%; Pred. No. 2.5e-23;
Matches 166; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 27 AAATCCATCTTGGGTGTGGGTGAGCACACCAAGAAATCTCCCTCAGTGGTTTGA 86
DB 21845 AATGTGCTTAGTGTATGAGAGACACACAAAGAGTTTTCGCTGATGTTTGT 21786
QY 87 GCTGTCCATCCGCTGTTCATTCATTTGGAATGCTGAGGAAATCTGCAACATGCCCAAG 146
DB 21785 GCTGTCCATGCTGTGCTGCTTTTATAGCAATGCTTAGGAAATCTGTGTGATGCCCAA 21726
QY 147 ACTGCCATGCAATCACATAGGAGCTTCCATTTGCTGAGCAATCGGTCGAGGGCG 206
DB 21725 ACAGCATGCTTTGACCATTTGAGCTTCATCTTGGGACAGGTGTTGGTCAAGAGCT 21666
QY 207 GAGCGCATTTGCTCTGAAGGCACTGGCTGCAAGAGCGACCTGATTCACACCGCTGGCT 266
DB 21665 GAGCGTTACCGTCTCAAGCAGTAGCTGAGAAATGCTTCTGTACCGCTATGTTTCT 21606
QY 267 G 267
DB 21605 G 21605

RESTART 10
AX654660 1104 bp DNA linear PAT 22-MAR-2003
LOCUS AX654660
DEFINITION Sequence 4530 from Patent WO03000898.
ACCESSION AX654660
VERSION AX654660.1 GI:29157474
KEYWORDS
SOURCE Oryza sativa
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzoideae; Oryzae; Oryza.
1
Chang, H.S., Chen, W., Cooper, B., Glazebrook, J., Goff, S.A., Hou, Y.M.,
Katagiri, F., Quan, S., Tao, Y., Whitham, S., Xie, Z., Zhu, T. and Zou, G.
Plant genes involved in defense against pathogens
Patent: WO 03000898-A 4530 03-JAN-2003;
Syngenta Participations AG (CH)
FEATURES
source
1..1104
/organism="Oryza sativa"
/mol_type="genomic DNA"
/db_xref="taxon:4530"
BASE COUNT 278 a 268 c 281 g 274 t 3 others
ORIGIN
Query Match 17.9%; Score 120.6; DB 6; Length 1104;
Best Local Similarity 76.3%; Pred. No. 2.1e-23;
Matches 161; Conservative 0; Mismatches 49; Indels 1; Gaps 1;

QY 15 TCGGATGCGCAGCAAC-TCCATCTTGGGTGTGGGTGAGCACACCAAGAAATCTCCCC 73
DB 729 TGGTATGCGACGACGTTCCCTTGGCATCTGGAGGAGCACACGAGAAATCTCAGT 788
QY 74 TCAGTGGTTGAGCTGCTCCATGCGCTGTTCATTCATTCGATGCTGAGGAAATCTGT 133
DB 789 GCAGTGGTTGCGCAGTCCATCGCGGTACCTTTCATAGGCGATGCTCAGAACTCTGT 848
QY 134 CAACATGCCCAAGACTGCGATGCGATTCACCATAGCAGCTCCATTTGTTGTCAGACAA 193

Db 849 GCTGATGCCAAAGACAGCCATGCGCTTACCATAGCTGCTCAATATTGGGTACAGACAAT 908

QY 194 CGGGTCGAGCGGAGCGCATTCGTCGAAG 224

Db 909 TGGTTCGAGAGCTGAGCGTATCAGATTGAAG 939

RESULT 11

AE017123/c

LOCUS

DEFINITION

Oriza sativa (japonica cultivar-group) chromosome 10, section 77 of the complete sequence.

ACCESSION

AE017123 AE018959

VERSION

AE017123.1 GI:31433699

KEYWORDS

Oriza sativa (japonica cultivar-group)

SOURCE

Oriza sativa (japonica cultivar-group)

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE

1 (bases 1 to 95893)

AUTHORS

The Rice Chromosome 10 Sequencing Consortium

CONSRMT

IN-depth view of structure, activity, and evolution of rice chromosome 10

TITLE

Science 300, 1566-1569 (2003)

JOURNAL

2 (bases 1 to 95893)

AUTHORS

Buell, C.R., Wang, R.A., McCombie, W.R., Messing, J. and Yuan, Q.

TITLE

Direct Submission

JOURNAL

Submitted (05-MAY-2003) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

COMMENT

This is the pseudomolecule for rice chromosome 10, which was constructed by resolving discrepancies between overlapping BACs, trimming the overlap regions, and linking the unique sequences to form a contiguous sequence. Genes in individual BAC clone were identified by a combination of several methods: Gene prediction programs, searches of the complete sequence against a peptide database and EST databases. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are annotated by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>).

FEATURES

source

1..95893

Location/Qualifiers

/organism="Oryza sativa (japonica cultivar-group)"

/mol_type="genomic DNA"

/cultivar="Nipponbare"

/db_xref="taxon:39947"

/chromosome="10"

misc_feature

1..95317

/note="Chromosome Sequence Derivation: nucleotide sequence in this region was derived from BAC clone OSUNBa0056G17 (GB:AC018727)."

gene

complement(<1834..>3425)

/locus_tag="OSUNBa0056G17.12"

/notes="EST D15631, C28081, C93504 from this gene"

mrna

complement(join(<1834..1923,2955..3039,3122..>3425))

/locus_tag="OSUNBa0056G17.12"

cds

complement(join(1834..1923,2955..3039,3122..3417))

/locus_tag="OSUNBa0056G17.12"

/codon_start=1

/product="unknown protein"

/protein_id="AAP55179.1"

/db_xref="GI:31433700"

/translation="MRWCAAAGLGLGRLLSSVAPAHRLPPLPFPSS RHLLPLGLHPSPPMQVQVHFHFAAKDRAPRTPTTSKVYKYIKLPSSMK FRFVNMVGGVRRWAGKRRHNAHLKSKAKRRLKPAVLHAYAKYIKLNPCS"

<4368..>7204

/locus_tag="OSUNBa0056G17.8"

/note="Similar to peptide transporter AtPTR2 B GB:AAB00858

GI:633940 (Arabidopsis thaliana); EST AU068484 from this gene"

join(<4368..4453,4697..4763,5040..5257,5412..5968,6385..>7204)

/locus_tag="OSUNBa0056G17.8"

join(<4406..4453,4697..4763,5040..5257,5412..5968,6385..>7204)

/locus_tag="OSUNBa0056G17.8"

/codon_start=1

/product="putative peptide transporter"

/protein_id="AAP55180.1"

/db_xref="GI:31433701"

/translation="MGWESGDAQLLPHHOASNSOHTKPPFNKAPALILAFEFLE SIAYAGISLNLVYLGTVLHGTTASNAAYVOTWNGTTELPVLGAKFLATYNGYKTI AISTVFLIGLLVITASVAPISQCPAPCNSSCPATGQTFVFFTLALYLSVGTGV KALLPFGADQYNSLNEESKKQSFSLFIANLGLVFGTGVVWIQQNVANSLGF GISSICIVATVAFVLPVQLPSPGSLFVYVFAVFKGKLEVPADNALLG DDADLNGOSVLAHTDGRFLDKAAVFEVEEINKDEGGGGLQCSQVVEEVK ILLMLPIWTVSLVLAASLGTTATTFVQGNAMTKISFSVPAASINSAEVI PMFWV VFDOTVVPPIARTYTCNPAGLTOLQRMGVRLLAVPALAAVAALVETWRLRSVRDGN LSIAMQLPQVILLACSDVFCGIAQLEFFVSEAPVMSLCSAFSLALSLGYVNSLV VSIYAVVTTYSNGKWLPAIDLNDGHDYFVLTWLTGISAINFVYVAAFAKNYTVKRLAVP HSH"

complement(<7577..>12536)

/locus_tag="OSUNBa0056G17.19"

/note="similar to polyprotein GB:AAD13304 GI:4235644 (Lycopodium obscurum)"

complement(join(<7577..7689,7783..7959,8059..8184,10478..10687,10865..10995,11095..11252,11633..11691,12215..>12536))

/locus_tag="OSUNBa0056G17.19"

complement(join(7577..7689,7783..7959,8059..8184,10478..10687,10865..10995,11095..11252,11633..11691,12215..>12536))

/locus_tag="OSUNBa0056G17.19"

/codon_start=1

/product="putative polyprotein"

/protein_id="AAP55181.1"

/db_xref="GI:31433702"

/translation="MDLDAKFDLLFKTIADNEKRMABEERTRADYLELKKTVGGRIP AMEKVSELGSLQSLTKVELEGMTROVKFPFPAASAGVGGAKKEQFGVAG MFTNGKINIQASHYFKRLKAWGTGESSRSLIGHWQCTELMLLIDSGTHS FMDISGKMGVSPFLKRLNLLNGDIIIGDMWLQSPFPMQVDMQSKMWEIQNNQ PRLQEPTELPKRYCDHSIPVLPVATPVLNLRPKFPAKLDIEIKQISMLKAGVLIQ PSHAPSPALLVKKDVFHVSUKSAVSGSLVQSLPTDITSLTFLPHLDRLRLK KAALWQANFQGGELVKCDPEEDKDRREPSIEKAKEAGRAEDLGLSFSSEBAQ YQDIELAIGVGSGSGSILLDLRGNVSHYVAKLCVCLSESL"

<14035..>15652

/locus_tag="OSUNBa0056G17.26"

/note="predicted by genscan"

join(<14035..14164,15507..>15652)

/locus_tag="OSUNBa0056G17.26"

join(14035..14164,15507..15652)

/locus_tag="OSUNBa0056G17.26"

/codon_start=1

/product="hypothetical protein"

/protein_id="AAP55182.1"

/db_xref="GI:31433703"

/translation="MVNTMPLPPAPPBILLYPIEDQRKAGVRRKDGELHGRSDGHEG TPGEAGWDECNPVKQCSITKYASRGQCGGLGIGQNGELNGLQ"

<18839..>20638

/locus_tag="OSUNBa0056G17.27"

/note="similar to peptide transporter AtPTR2 B GB:AAB00858 GI:633940 (Arabidopsis thaliana)"

join(<18839..19159,19244..>20638)

/locus_tag="OSUNBa0056G17.27"

join(18839..19159,19244..20638)

/locus_tag="OSUNBa0056G17.27"

/codon_start=1

/product="putative peptide transporter"

/protein_id="AAP55183.1"

/db_xref="GI:31433704"

/translation="MSNNMMGSPSPSRGRDEKKRWKWLGPATILGPELLSIAF

SQVALNLVYLVLTAVLHGTAFNAHVDTWNGTTFIVPVIAPLADSYMGKRYTILASL
 LPYLAVGLVLTAVSAPVLSLPAPCTGVPSCPATGTQFSVFPLALVLSIGTGGVKSAL
 LPPGAQVQERDDHDDQSGAPAPKTKQSPFSWFGAINLGIFFVAGTILSVWQGVNSW
 ALGFGIATCLLIAGAAATPCTXRYRPLTGDTPFKALILRVLVAAEFNRRTKPLPDA
 DDGLYEVDDKNGGDDKLAHTEGRLWDKAAVRVDGGGVNEVCTSEVERKVL
 ARIVPIWVPLVLAASLQGTTFITQQGMAMDTLVFGGRFVPVAVLSVSEVPMFLM
 VLLHDVMPVAVRNRSGGLTQLQRMGVGRVLVYVAMATAALVERRLRGRKRSNI
 LWQVPQVVLVLAGSVFSGIADLEFFYGAPGSMRSICSAFSLALSIGFYVNSLVVTI
 VAAVTKRSGDGSGLAPLDLTAHLDYFYLWALISLANLALYLLAARYKSKPSPPH
 SSSH"

gene
 /locus_tag="OSJNBa0056G17.28"
 /note="similar to Mutator protein MUDRA GS:AAA81535
 GI:595816 (Zea mays)"
 mRNA
 Join(21897..23030,23238..24577,24883..24955,25538..25808,
 25903..26187,26501..26568,26661..26866,26961..27255,
 27337..27435,27513..27870,28174..28415)
 /locus_tag="OSJNBa0056G17.28"
 CDS
 Join(21897..23030,23238..24577,24883..24955,25538..25808,
 25903..26187,26501..26568,26661..26866,26961..27255,
 27337..27435,27513..27870,28174..28415)
 /locus_tag="OSJNBa0056G17.28"
 /codon_start=1
 /product="putative Mutator protein"
 /protein_id="AAP55184.1"
 /db_xref="GI:31433706"

/translation="MTVQLVFSVPTPEELMSRIKHELDNSDSVGMQGRYDVGGV
 MSHKMLDNGEIEWTQYIDIVGSHFSLVFAWKDGRIGKELIDNLSLIESP
 INCHLSERCKKLEEDVEGEGVNDMAVDEVBREEDVREVAHEAVPREEV
 HAEDYVREKVENLEEDVEGEGVNDMAVDEVBREEDVREVAHEAVPREEV
 KFYTEBEAMIRSGLPWLEDTDEDEASDSDDRDLVRKMSHEKTVAKLV
 GRNPITOPENTRSLGATADQPOVDGAFPMCDPRKGLFRSMDLKLWLOSYSI
 RVHRPVHKESNASKYTVACIDRCKQINARKSGDRMVRVINGNDDEAVRPL
 TLMRAIKAKPTMFEVNAHPEKSRMVDGQVRVGRAYWIFGGSIEAFKHLRVLAL
 DGTFTGTCGLTALGADGLHLVLAFLAVEKENTSNWFNMLNKLGNRE
 VCIIEDRPIGLINSIHHMPLHTIHHRCWHPHCANFYTAGATDQKDLERCOIN
 EKAPFLDKLGMVGRPKKLEHMDPLVKVWARFDTNGRHSIMTSMAFSFN
 VLRGKLPVTAIVATVFSKNSVWDEHKEATVILSKKWTXKMDLKEEQRTIL
 GORACFOFPKMYEVSQGVTAAGVWGRHYVVAARDNCSOPPOLHLPCHM
 ITVCKRLGLDVRMAVEASNAKAVQSPSPPEYLDPSOMPSPYGEFVPLSLK
 NNTGRRLRTRKNDKAYKGSANRMEHPGLDFYEEKRAPQIASGRILKLRV
 RGHATHILFDVPILOKALLAVTWAQRPVLRVNAALVLRVNAALVLRVNAALVLRV
 GELTVLVEDVAMILGTRQAVTGTASGNRVERVEYIGLEFPVAPDQGRKTS
 CGLTSMWLNFOQPAEDATVORYCAYLYIFGSLFFDSGDGMMWMLFLAD
 MDEAVVGRNLAHYTKEMDYLQHEVWMPYQAEALBELNPMCHIEDIALTKRC
 PLICYAVFQWCHRMVQFGLQTPHFRFSTIDLHKVRKKNKVTMDYHQDHI
 TWKPEFENGVDQGNHTFEDLYLWHRPLVLRPAWTLADIAADPDVEONE
 YDTRLTGTVTGPVRVARELLRTVNDAGVALTAPGSEGGTLRNALQRLOR
 CKLAARLGRSTDVVEHAHAHOEGDEGEVDEGEQDKAEBQDEBEEBEEB
 GDEBEAELGSLQEDAPESQPLSQPSQRRRWPQDMRYTPDVRPRIRAP
 VHALSSQSPSHPEGLGWMATKSLHSARFVGTLEPTVSAIMAGAEVPAKPSRRL
 AGLDLGRQGVPPQAGWR"
 complement(31524..31948)
 /locus_tag="OSJNBa0056G17.20"
 /note="predicted by genemark hm"

complement(join(31524..31652,31754..31948))
 /locus_tag="OSJNBa0056G17.20"
 complement(join(31524..31652,31754..31948))
 /locus_tag="OSJNBa0056G17.20"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AAP55185.1"
 /db_xref="GI:31433706"

Query Match 17.9%; Score 120.6; DB 8; Length 95893;
 Best Local Similarity 76.3%; Pred. No. 3.2e-23;
 Matches 161; Conservative 0; Mismatches 49; Indels 1; Gaps 1;

QY 15 TCGATGGCAGCAAC-TCCATCTTGGGCTGGCGTACGACACCAAGAAATTCCTCCC 73
 DB 82603 TGTATGGCAGCAACGTTCCCTTGGCATCTGGAGGAGCAGGAGAAATTCCTAGT 82544
 QY 74 TCAGTGGTTGACGTGTTCATGCCCGCTTCATTCATTGGAGTCTGAGGAATCTGT 133

Db 82543 GCAGTGGTTGGCGAGTCCATGGCGGTACCTTTTCATAGGCATGCTCAGGAAGTCTGT 82484
 QY 134 CAACATGCCCAACATGCCATGCCATTCACCTAGCAGCCTCCTATTGTTGGTCAGACAAT 193
 Db 82483 GCTGATGCCAAGACAGCAGCCATGGCGTTCACCTAGTGCCTCAATATTGGTTCAGACAAT 82424
 QY 194 CGGTCGAGCGCGAGCGCATTCGTCTGAAG 224
 Db 82423 TGGTTCGAGAGCTGAGCGTATCAGATTGAAG 82393

RESULT 12
 AC018727/c
 LOCUS
 DEFINITION
 Oryza sativa chromosome 10 BAC OSJNBa0056G17 genomic sequence,
 complete sequence.
 ACCESSION
 AC018727
 VERSION
 AC018727.10 GI:12039362
 KEYWORDS
 HTG.
 ORGANISM
 Oryza sativa (japonica cultivar-group)
 Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzoideae; Oryza.
 REFERENCE
 1 (bases 1 to 139999)
 Buell, C.R., Yuan, Q., Moffat, K.S., Hill, J.N., Burr, P.C., Hsiao, J.,
 Ziemann, V., Pa, G., Bowman, C.L., Fujii, C.Y., VanAken, S.E.,
 Bowman, C.L., Craven, B., Utterback, T.R., Khalak, H., Feldblyum, T.V.,
 Quackenbush, J., White, O., Salzberg, S.L., and Fraser, C.M.
 Oryza sativa chromosome 10 BAC OSJNBa0056G17 genomic sequence
 Unpublished
 2 (bases 1 to 139999)
 Buell, R.
 Direct Submission
 Submitted (17-DEC-1999) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 3 (bases 1 to 139999)
 Buell, R.
 Direct Submission
 Submitted (05-JAN-2001) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA, rbuell@tigr.org
 4 (bases 1 to 139999)
 Buell, R.
 Direct Submission
 Submitted (13-MAR-2001) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA, rbuell@tigr.org
 5 (bases 1 to 139999)
 Buell, R.
 Direct Submission
 Submitted (29-AUG-2001) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA, rbuell@tigr.org
 On Jan 5, 2001 this sequence version replaced gi:12025621.
 Address all correspondence to:rice@tigr.org

BAC clone OSJNBa0056G17 is from Oryza sativa chromosome 10.
 The orientation of the sequence is from SP6 to 3' end of the BAC
 clone.
 Genes were identified by a combination of several methods: Gene
 prediction programs including GENSCAN and GENSCAN+ (Chris Burge,
 http://CCR-081.mit.edu/GENSCAN.html), GeneMark-ES (Mark Borodovsky,
 http://genemark.biology.gatech.edu/GeneMark/), Egenes
 (http://www.softberry.com/), and GeneSplicer (Mihaila Pertea and
 Steven Salzberg, contact mpertea@tigr.org), searches of the
 complete sequence against a peptide database and the plant EST
 database at TIGR (http://www.tigr.org/tdb/tgi.shtml). Annotated
 genes are named to indicate the level of evidence for their
 annotation. Genes with similarity to other proteins are named after
 the database hits. Genes without significant peptide similarity but
 with EST similarity are named as unknown proteins. Genes without
 protein or EST similarity, that are predicted by more than two gene
 prediction programs over most of their length are annotated as
 hypothetical proteins. Genes encoding tRNAs are predicted by

source
1. .1035
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
BASE COUNT 282 a 235 c 247 g 271 t
ORIGIN

Query Match 15.6%; Score 105.2; DB 6; Length 1035;
Best Local Similarity 68.2%; Pred. No. 5.8e-19;
Matches 146; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
QY 27 AAATCCATCTGGGGTGTGGCGTGAGCACACCAAGAAATTCCTCCCTCAGTGGTTTGA 86
DB 652 AATGTGCTTTAGGGTCTGGAGAGACACACTGAAAGTTTTCGGCATCTTGGTTTATA 711
QY 87 GCTGTCCATCCCGCTGTTCATTCAATGGAATCTGAGAAATCTGTCAACATGCCCAAG 146
DB 712 GCTCTTCACGACGGTTCATTATAGGAATCTGAGAAAGTCAGTGTGATGCTTAA 771
QY 147 ACTGCCATGGCATTCCATAGCAGCTCCATCTTGGTCAGACATCCGGTCGAGGGCG 206
DB 772 ACAGCTATGGTATTCACCATAGCAGCATCAGTTTGGGACAGGTGATTGGGTCAAGACA 831
QY 207 GAGCGATTCTGTGAAGGCACTGGTCAAGA 240
DB 832 GAGAGCGGAGACTCAAGTCTGTAGCTGAGAAGA 865

RESULT 15
AX651340
LOCUS AX651340 1035 bp DNA linear PAT 22-MAR-2003
DEFINITION Sequence 128 from Patent WO03000898.
ACCESSION AX651340
VERSION AX651340.1 GI:29154158
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
AUTHORS Chang, H.S., Chen, W., Cooper, B., Glazebrook, J., Goff, S.A., Hou, Y.M.,
Katagiri, F., Quan, S., Tap, Y., Whitham, S., Xie, Z., Zhu, F. and Zou, G.
TITLE Plant genes involved in defense against pathogens
JOURNAL Patent: WO 03000898-A 128 03-JAN-2003;
SYNOPSIS Syngenta Participations AG (CH)
FEATURES
source
1. .1035
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
BASE COUNT 282 a 235 c 247 g 271 t
ORIGIN

Query Match 15.6%; Score 105.2; DB 6; Length 1035;
Best Local Similarity 68.2%; Pred. No. 5.8e-19;
Matches 146; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
QY 27 AAATCCATCTGGGGTGTGGCGTGAGCACACCAAGAAATTCCTCCCTCAGTGGTTTGA 86
DB 652 AATGTGCTTTAGGGTCTGGAGAGACACACTGAAAGTTTTCGGCATCTTGGTTTATA 711
QY 87 GCTGTCCATCCCGCTGTTCATTCAATGGAATCTGAGAAATCTGTCAACATGCCCAAG 146
DB 712 GCTCTTCACGACGGTTCATTATAGGAATCTGAGAAAGTCAGTGTGATGCTTAA 771
QY 147 ACTGCCATGGCATTCCATAGCAGCTCCATCTTGGTCAGACATCCGGTCGAGGGCG 206
DB 772 ACAGCTATGGTATTCACCATAGCAGCATCAGTTTGGGACAGGTGATTGGGTCAAGACA 831
QY 207 GAGCGATTCTGTGAAGGCACTGGTCAAGA 240
DB 832 GAGAGCGGAGACTCAAGTCTGTAGCTGAGAAGA 865

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2003, 10:08:18 ; Search time 1929 Seconds
(without alignments)
8479.478 Million cell updates/sec

Title: US-09-868-025-1

Perfect score: 673

Sequence: 1 ttatcctgctcgtcgtgat.....actgtgtgtaaaaaaaaaa 673

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 22781392 seqs, 12152238056 residues

Word size : 0

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estnu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_htc.*
9: gb_est1.*
10: gb_est2.*
11: gb_hcc.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_pln.*
20: em_gss_vrt.*
21: em_gss_fun.*
22: em_gss_mam.*
23: em_gss_mus.*
24: em_gss_pro.*
25: em_gss_rod.*
26: em_gss_phg.*
27: em_gss_vrl.*
28: gb_gsel.*
29: gb_gsel2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	528	78.5	723	9 AU068433	AU068433 AU068433
2	528	78.5	776	14 CB000367	CB000367 S345U H05
3	526	78.2	805	14 CB660999	CB660999 OSJNE03E
4	509	75.6	712	9 AU075373	AU075373 AU075373

C 5	493	73.3	780	14	CB655151
C 6	492	73.1	727	14	CB655085
C 7	475	70.6	763	14	CB663774
C 8	467	69.4	705	9	AU165747
9	466	69.2	718	9	AU031690
C 10	463	68.8	850	14	CB619892
C 11	443	65.8	792	14	CB647260
C 12	442	65.7	682	14	CB647623
C 13	442	65.7	777	14	CB627453
C 14	441	65.5	799	14	CB620522
C 15	441	65.5	836	14	CB659692
C 16	440	65.4	622	13	BQ068858
C 17	390	57.9	816	14	CB633818
C 18	358	54.7	486	9	AT003421
C 19	339	50.4	653	14	CB643278
C 20	312	46.4	510	12	BM419698
C 21	243	36.1	768	14	CB649882
C 22	234	34.8	691	10	BE229010
C 23	227	33.7	382	9	AU030190
C 24	215	31.9	378	9	AU093781
C 25	204	30.3	504	12	BI811292
C 26	166	24.7	449	9	AT003707
C 27	161	23.9	321	10	BE230720
C 28	144	21.4	345	14	D23912
C 29	137	20.4	190	12	BI811260
C 30	137	20.4	514	12	BI809607
C 31	130	19.3	782	28	AZ133549
C 32	111	16.2	479	12	BI809674
C 33	109	16.2	693	14	CB655084
C 34	109	16.2	726	14	CB655150
C 35	107	15.9	410	12	BM420265
C 36	101	15.0	275	12	BI809342
C 37	95	14.1	419	14	D48832
C 38	90	13.4	464	9	AU184049
C 39	87	12.9	149	14	CA765737
C 40	83	12.3	681	10	BF430583
C 41	82	12.2	416	14	C73513
C 42	77	11.4	459	9	AT003640
C 43	73	10.8	898	14	CB660998
C 44	65	9.7	484	12	BI797263
C 45	50	7.4	427	29	AG021857

ALIGNMENTS

RESULT 1
AU068433
LOCUS
DEFINITION AU068433 Rice callus Oryza sativa (japonica cultivar-group) cDNA
clone C30250_52, mRNA sequence.
ACCESSION AU068433
VERSION AU068433.1 GI:5003284
KEYWORDS EST.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
REFERENCE Yamanoto, K. and Sasaki, T.
AUTHORS Rice cDNA from callus (1998)
TITLE Unpublished
JOURNAL
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsasaki@abr.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/
PROJECT = 'RGP'
FEATURES Location/Qualifiers
1..723

723 bp mRNA linear EST 01-APR-2002
AU068433 Rice callus Oryza sativa (japonica cultivar-group) cDNA
clone C30250_52, mRNA sequence.
ACCESSION AU068433
VERSION AU068433.1 GI:5003284
KEYWORDS EST.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
REFERENCE Yamanoto, K. and Sasaki, T.
AUTHORS Rice cDNA from callus (1998)
TITLE Unpublished
JOURNAL
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsasaki@abr.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/
PROJECT = 'RGP'
FEATURES Location/Qualifiers
1..723

/organism="Oryza sativa (japonica cultivar-group)"

/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="C30250.52"
/clone_lib="rice callus"
/note="vector: pBluescript II SK-; Site 1: SalI; Site 2:
NotI; cDNA prepared from rice callus mRNAs by using
oligo(dG) as a primer and ligating to the SalI-NotI site
of pBluescript II SK+ phagemid."
BASE COUNT 183 a 154 c 183 g 203 t

Query Match 78 5%; Score 528; DB 9; Length 723;
Best Local Similarity 99.7%; Pred. No. 5.8e-173;
Matches 628; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 35 TCTTGGGTGGCGTGAGCACACCAAGAAATTCCTCCCTCAGTGGTTTCAGCTGCCA 94
DB 62 TCTTGGGTGGCGTGAGCACACCAAGAAATTCCTCCCTCAGTGGTTTCAGCTGCCA 121
QY 95 TGCCGCTGTTCCATTGGAAGCTGAGCAATCTGTCAACATGCCAGACTGCCAT 154
DB 122 TGCCGCTGTTCCATTGGAAGCTGAGCAATCTGTCAACATGCCAGACTGCCAT 181
QY 155 GGCATTCCACATAGCAGCTCCATTGTTGTGTCAGCAATCGGTCGAGGCGGAGCGCAT 214
DB 182 GGCATTCCACATAGCAGCTCCATTGTTGTGTCAGCAATCGGTCGAGGCGGAGCGCAT 241
QY 215 TCGTCTGAAGGCACCTGCTCAAGAGCGAGCGCTGATTCACACCGTGTGCATGTGA 274
DB 242 TCGTCTGAAGGCACCTGCTCAAGAGCGAGCGCTGATTCACACCGTGTGCATGTGA 301
QY 275 TCCAAACAAGACTGCAAAATTCAGTGACACCGAGGCGAAGCGCATCGCTCGCGAT 334
DB 302 TCCAAACAAGACTGCAAAATTCAGTGACACCGAGGCGAAGCGCATCGCTCGCGAT 361
QY 335 GAGATGATCGGGGAGCGGCTTCTGCTGGTGTGCTGCTCCACACCAAGCATGTGTTT 394
DB 362 GAGATGATCGGGGAGCGGCTTCTGCTGGTGTGCTGCTCCACACCAAGCATGTGTTT 421
QY 395 CTGATTGCTCACTGATTGGAAAATTTGATCTACAGTATCCCTGGAGAGTGGAGATTG 454
DB 422 CTGATTGCTCACTGATTGGAAAATTTGATCTACAGTATCCCTGGAGAGTGGAGATTG 481
QY 455 ATATTGAGTCTATTTATCTGTGATGATTAATGCTGCTGCTCCCTCAGAAATTCG 514
DB 482 ATATTGAGTCTATTTATCTGTGATGATTAATGCTGCTGCTCCCTCAGAAATTCG 541
QY 515 TTTGTTTGTGGGATGAGCAAGTGGAAATAGAGTGTCTACTATATACAGCATCTCTGTT 574
DB 542 TTTGTTTGTGGGATGAGCAAGTGGAAATAGAGTGTCTACTATATACAGCATCTCTGTT 601
QY 575 GTTAAGTTTGCAGTTCTGAGTTCATGTATCTGTAAATTTGATGATGCTGGATTTCTACT 634
DB 602 GTTAAGTTTGCAGTTCTGAGTTCATGTATCTGTAAATTTGATGATGCTGGATTTCTACT 661
QY 635 ATTTATCAATCGTCATTATATCTGTGTGTA 664
DB 662 ATTTATCAATCGTCATTATCTGTGTGTA 691

```

RESULT 2
CB000367
LOCUS
DEFINITION CB000367 776 bp mRNA linear EST 10-JAN-2003
(japonica cultivar-group) cDNA library Oryza sativa
ACCESSION (japonica cultivar-group) cDNA 5', mRNA sequence.
VERSION CB000367.1 GI:27577672
KEYWORDS EST.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE AUTHORS

de los Reyes, B. G., Morsy, M., Gibbons, J., Varma, T.S.N., Antoine, W.,
Redus, M., McGath, J.M. and Halgren, R.
Development of a chilling stress EST library of germinating rice
(Oryza sativa L. subsp. japonica) enriched with stress-related and
novel genes
Unpublished

JOURNAL COMMENT

Contact: Benildo G. de los Reyes
Plant Genomics Lab., Department of Crop, Soil and Environmental
Sciences

University of Arkansas
115 Plant Science Building, Fayetteville, AR 72701, USA

Tel: (479)-575-7465

Fax: (479)-575-8435

Email: breyes@uark.edu

Plate: S345U row: H column: 05

Seq primer: T3.

FEATURES source

1..776
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="PI560247"
/db_xref="taxon:39947"
/tissue_type="coleoptile, radicle, prophyll, immature leaf"
/dev_stage="embryo at different stages of germination and
shoots of germinated seeds under chilling stress
(130C/100C)"
/lab_host="Soir"
/clone_lib="Rice cold stress germination cDNA library"
/note="Organ: seedlings; Vector: Lambda Uni-Zap XR excised
as pBluescript; Site 1: EcoRI; Site 2: XhoI; The cDNAs
were derived from reverse transcription of mRNA samples
from seeds at different stages of germination and
seedlings at early phase of growth under chilling stress
(130C/100C). The mRNA pool was used as template for double
stranded cDNA synthesis using the Stratagene Uni-Zap XR
cDNA synthesis and library kit. A total of 150,000 phages
were excised from the primary library as pBluescript
phagemid clones. Enrichment of the primary excised library
with chilling-induced transcripts was performed by
hybridizing the primary excised library colony lifts with
the PCR-select subtraction product, with cold germinated
cDNA as tester and control temperature-germinated cDNA as
driver."

BASE COUNT 170 a 180 c 203 g 223 t

ORIGIN

Query Match 78 5%; Score 528; DB 14; Length 776;
Best Local Similarity 99.7%; Pred. No. 5.6e-173;
Matches 628; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 35 TCTTGGGTGGCGTGAGCACACCAAGAAATTCCTCCCTCAGTGGTTTCAGCTGCCA 94
DB 137 TCTTGGGTGGCGTGAGCACACCAAGAAATTCCTCCCTCAGTGGTTTCAGCTGCCA 196
QY 95 TGCCGCTGTTCCATTGGAAGCTGAGCAATCTGTCAACATGCCAGACTGCCAT 154
DB 197 TGCCGCTGTTCCATTGGAAGCTGAGCAATCTGTCAACATGCCAGACTGCCAT 256
QY 155 GGCATTCCACATAGCAGCTCCATTGTTGTGTCAGCAATCGGTCGAGGCGGAGCGCAT 214
DB 257 GGCATTCCACATAGCAGCTCCATTGTTGTGTCAGCAATCGGTCGAGGCGGAGCGCAT 316
QY 215 TCGTCTGAAGGCACCTGCTGCAAGAGCGAGCGTGTATTCACCGCTGGCTGACATGTA 274
DB 317 TCGTCTGAAGGCACCTGCTGCAAGAGCGAGCGTGTATTCACCGCTGGCTGACATGTA 376
QY 275 TCCAAACAAGACTGCAAAATTTGAGTGACACCGAGGCGAAGCATGGGATCCGCTCGCGAT 334
DB 377 TCCAAACAAGACTGCAAAATTTGAGTGACACCGAGGCGAAGCATGGGATCCGCTCGCGAT 436

```

Qy	335	GAAGATGATGGGGGACGGGGCTTCTGGTGGTCTGCTCTCAACACCAAGCATGTGTTT	394
Db	437	GAAGATGATGGGGGACGGGGCTTCTGGTGGTCTGCTCTCAACACCAAGCATGTGTTT	496
Qy	395	CTGATTGCTCACTGATTGGAAAAATTTGTATCTACACAGTATCCTCGAGAGTGGAGATTG	454
Db	497	CTGATTGCTCACTGATTGGAAAAATTTGTATCTACACAGTATCCTCGAGAGTGGAGATTG	556
Qy	455	ATATTGAGTCTATTTTATCTTTGTGATGTAAATTCCTTTGCTTGTCCTCAGAAGTATTGCG	514
Db	557	ATATTGAGTCTATTTTATCTTTGTGATGTAAATTCCTTTGCTTGTCCTCAGAAGTATTGCG	616
Qy	515	TTTGTGTTTGGGATGAGCAAGTGGAAATTAAGAGTGCTACTATATACAGCATCAATCTCTTT	574
Db	617	TTTGTGTTTGGGGTGAGCAAGTGGAAATTAAGAGTGCTACTATATACAGCATCAATCTCTTT	676
Qy	575	GTTTAAGTTTGGCAGTCTCGCAGTTTCATCTACTGTAAATTTGATGATGTGGATTTCTTACT	634
Db	677	GTTTAAGTTTGGCAGTCTCGCAGTTTCATCTACTGTAAATTTGATGATGTGGATTTCTTACT	736
Qy	635	ATTATATCAATCGTCATTATACCTGTGTGTA	664
Db	737	ATTATATCAATCGTCATTATACCTGTGTGTA	766

RESULT 3
CB6S0999/c
LOCUS
DEFINITION
OSJNE303E21.r OSJNED Oryza sativa (japonica cultivar-group) cDNA clone OSJNE303E21 3' mRNA sequence.
805 bp mRNA linear EST 09-APR-2003

RECORD NUMBER	CB660999.1	GI:29664724
VERSION	EST.	
KEYWORDS		
SOURCE	Oryza sativa (japonica cultivar-group)	
ORGANISM	Oryza sativa (japonica cultivar-group)	
	Eukaryota, Viridiplantae, Streptophyta, Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.	
REFERENCE	1 (bases 1 to 805)	
AUTHORS	Jantasuriyarat, C., Lu, G., Gowda, M., Hatfield, J., Zhou, B., Mazur, E., Kudrna, D., Dean, R., Soderlund, C., Wing, R. and Wang, G.	
TITLE	Large-scale identification of ESTs involved in the interaction between rice and Magnaporthe grisea	
JOURNAL	Unpublished	

CONTACT: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088. USA

Tel: 520 626 3967
Fax: 520 621 9288
Email: <http://genome.arizona.edu>
PCR Primers
FORWARD: gta aaa cga cgg cca ctg
BACKWARD: gga aac agc tat gac cat g
Plate: 03 row: E column: 21
Seq primer: gga aac agc tat gac cat g

```

FEATURES
    seq primrpt="99a and 99b cat gac cat g".
    Location/Qualifiers
        1..805
            /organism="Oryza sativa (japonica cultivar-group)"
            /mol_type="mRNA"
            /cultivar="Nipponbare"
            /db_xref="taxon:39947"
            /clone="OSJNEd03521"
            /tissue_type="Leaf"
            /dev_stage="3 week"
            /lab_host="DH10B"
            /clone_lib="OSJNEd"

```

BASE COUNT	ORIGIN
226 a	212 c 190 g 177 t
ANO1; 24 hrs after inoculation with Rice Blast (C9240-1)	

Query Match	78.2%	Score 526;	DB 14;	Length 805;
Best Local Similarity	99.7%	Pred. No. 2,7e-172;		
Matches 626;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;
QY	35	TCCTGGGGTGTGGCGTGAGCACACCAAGAAATCTCCCTCAGTGGTTTCAGCTGTCCCA	94	
Db	628	TCCTGGGGTGTGGCGTGAGCACACCAAGAAATCTCCCTCAGTGGTTTCAGCTGTCCCA	569	
QY	95	TGCGCGTGTCCATTCCATTGGAATGCTGAGGAATCTGTCAACAATGCCCAAGACTGCGCAT	154	
Db	568	TGCGCGTGTTCCTATTTCATTGGAATGCTGAGGAATCTGTCAACAATGCCCAAGACTGCGCAT	509	
QY	155	GGCATTACCATAGCAGCCCTCCATTGTGTGTGACAGAAATCGGTCGAGGCGGAGCGCAT	214	
Db	508	GGCATTACCATAGCAGCCCTCCATTGTGTGTGACAGAAATCGGTCGAGGCGGAGCGCAT	449	
QY	215	TCGCTCTGAAGGCACTGGCTGTCAAAGACGACGCTGATTCCACCAACCGTGCTGACATGTA	274	
Db	448	TCGCTCTGAAGGCACTGGCTGTCAAAGACGACGCTGATTCCACCAACCGTGCTGACATGTA	389	
QY	275	TCCAAACAAAGACTGCATAATTGCAGTGNACCGGAGGCGAAGGCATGGATTCGGCTCGCGCAT	334	
Db	388	TCCAAACAAAGACTGCAGAAATTCGAGTGCACACCGGAGGCGAAGGCATGGGATTCGGCTCGCGCAT	329	
QY	335	GAAGATGATGGCGGGAACGGGCTTCTGGTGGTGTGCTGTCTCCAAACCAAGCATGTGTTT	394	
Db	328	GAAGATGATGGCGGGAACGGGCTTCTGGTGGTGTGCTGTCTCCAAACCAAGCATGTGTTT	269	
QY	395	CTGATTGCTCACTGATTGGAAAAATTTGTATCTACCAAGTATCCCTGGAGAGTGAGAGTTG	454	
Db	268	CTGATTGCTCACTGATTGGAAAAATTTGTATCTACCAAGTATCCCTGGAGAGTGAGAGTTG	209	
QY	455	ATATTGAGTCTATTTTATCTGTGTGATTAATTCCTTTTGGCTGTCTCCCTCAGAAGTATTCG	514	
Db	208	ATATTGAGTCTATTTTATCTGTGTGATTAATTCCTTTTGGCTGTCTCCCTCAGAAGTATTCG	149	
QY	515	TTTGTGTGGGATAGACAAGTGGAAATTAAGAGTGCTACTATATACACGATCATTCGT	574	
Db	148	TTTGTGTGGGATAGACAAGTGGAAATTAAGAGTGCTACTATATACACGATCATTCGT	89	
QY	575	GTTAAGTTTCCCAAGTTCTGCAGTTCATGTATCTGTAATTTGATGATGCTGGATTTCTACT	634	
Db	88	GTTAAGTTTCCCAAGTTCTGCAGTTCATGTATCTGTAATTTGATGATGCTGGATTTCTACT	29	
QY	635	ATTATCAATCGTCATTATCTGTGT	662	
Db	28	ATTATCAATCGTCATTATCTGTGT	1	

RESULT 4	
LOCUS	AU0753373
DEFINITION	AU0753373 Rice panicle at flowering stage Oryza sativa [japonica cultivar-group] cDNA clone #2966_4L, mRNA sequence.
ACCESSION	AU0753373
VERSION	AU0753373.1 GI:5055994
KEYWORDS	EST.
SOURCE	Oryza sativa (japonica cultivar-group)
ORGANISM	Oryza sativa (japonica cultivar-group) Eukaryota; Viridiplantae; Streptophyta; Tracheophyta; Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae; Erbartoideae; Oryzaceae; Oryza. 1 (bases 1 to 712)
REFERENCE	Sasaki,T. and Yamamoto,K. Rice cDNA from panicle at flowering stage Unpublished
AUTHORS	Contact: Takuji Sasaki
TITLE	National Institute of Acrobiological Resources
JOURNAL	Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
COMMENT	305-8602, Japan Tel: 81-298-38-7441 Fax: 81-298-38-7468 Email: tsasak@abrr.afric.go.jp, URL:http://tsgp.dna.afric.go.jp/


```

PROJECT = 'RGP'
FEATURES
  source
    1. 712
    /organism="Oryza sativa (japonica cultivar-group)"
    /mol_type="mRNA"
    /cultivar="Nipponbare"
    /db_xref="taxon:39947"
    /clone_lib="E2906.42"
    /dev_stage="flowering stage"
    /notes="Organ: panicle; Rice cDNA from panicle at flowering stage"
  BASE COUNT    175 a   156 c   181 g   199 t
  ORIGIN
    75.8%; Score 509; DB 9; Length 712;
    Best Local Similarity 99.7%; Pred. No. 2.2e-166;
    Matches 609; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
  QY 63 AAATTCCTCCCTCAGTGGTTTGAGCTGTCATCGCGCTGTTCCATTCATTGGAAATGCTG 122
  Db 97 AAATTCCTCCCTCAGTGGTTTGAGCTGTCATCGCGCTGTTCCATTCATTGGAAATGCTG 156
  QY 123 AGAAATCTGTCAACATGCCCAAGACTGCCATGGCAATCCACATAGCAGCTCCATTGTT 182
  Db 157 AGAAATCTGTCAACATGCCCAAGACTGCCATGGCAATCCACATAGCAGCTCCATTGTT 216
  QY 183 GGTCAACAATCCGGTTCGAGCGGAGCGCATTCGTCTGAAGGCACTGGCTGCAAGAGC 242
  Db 217 GGTCAACAATCCGGTTCGAGCGGAGCGCATTCGTCTGAAGGCACTGGCTGCAAGAGC 276
  QY 243 GAGCTGATTCACACACCGTGGCTGACATGTATTCACAAAGACTGCCAATTCGAGTGAC 302
  Db 277 GAGCTGATTCACACACCGTGGCTGACATGTATTCACAAAGACTGCCAATTCGAGTGAC 336
  QY 303 ACCGAGGCAAGCATGGGATCCGCTCGCATGAAGATGATGGCGGAGCGGGCTTCGTGT 362
  Db 337 ACCGAGGCAAGCATGGGATCCGCTCGCATGAAGATGATGGCGGAGCGGGCTTCGTGT 396
  QY 363 GGTGCTGCTGCTCCACACCAAGCATGTGTTCTTGATTTGCTACTGATTTGGAATTTGT 422
  Db 397 GGTGCTGCTGCTCCACACCAAGCATGTGTTCTTGATTTGCTACTGATTTGGAATTTGT 456
  QY 423 ATCTACAGTATCCCTGGAGTGGAGTGTATATGATGCTATTTATTTATCTTGTGATGT 482
  Db 457 ATCTACAGTATCCCTGGAGTGGAGTGTATATGATGCTATTTATTTATCTTGTGATGT 516
  QY 483 AATTGCTTTGCTTGCTCCACCAAGTATTCGTTTGTGTTGGATGACACCAAGTGGAT 542
  Db 517 AATTGCTTTGCTTGCTCCACCAAGTATTCGTTTGTGTTGGATGACACCAAGTGGAT 576
  QY 543 AAGAGTCTACTATATACAGATCATTCCTGTTTAAAGTTTGCAGTTCTGTCAGTTTCATG 602
  Db 577 AAGAGTCTACTATATACAGATCATTCCTGTTTAAAGTTTGCAGTTCTGTCAGTTTCATG 636
  QY 603 TATCTGTAATTTGATGATGCTGATTTCTACTATTTATCAATCGTCATTTACTGTGT 662
  Db 637 TATCTGTAATTTGATGATGCTGATTTCTACTATTTATCAATCGTCATTTACTGTGT 696
  QY 663 AAAAAAAAAA 673
  Db 697 AAAAAAAAAA 707
  RESULT 5
  CB655151/c
  LOCUS
  DEFINITION
    OSJUNE0812.r OSJUNE Oryza sativa (japonica cultivar-group) cDNA
    clone OSJUNE0812.3', mRNA sequence.
  ACCESSION
    CB655151
  VERSION
    CB655151.1 GI:29658876
  KEYWORDS
    EST.
    Oryza sativa (japonica cultivar-group)

```

```

ORGANISM
  Oryza sativa (japonica cultivar-group)
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
  Ehrhartoideae; Oryzaceae; Oryza.
  1 (bases 1 to 780)
  Jantassuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
  Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
  Large-scale identification of ESTs involved in the interaction
  between rice and Magnaporthe grisea
  Unpublished
  JOURNAL
  COMMENT
    Contact: Rod Wing
    Arizona Genomics Institute
    University of Arizona
    Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
    85721-0088, USA
    Tel: 520 626 3967
    Fax: 520 621 9288
    Email: http://genome.arizona.edu
    PCR Primers
    FORWARD: gta aaa cga cgg cca gtc
    BACKWARD: gga aac agc tat gac cat g
    Plate: 08 row: E column: 12
    Seq primer: gga aac agc tat gac cat g.
  FEATURES
    source
      Location/Qualifiers
        1..780
        /organism="Oryza sativa (japonica cultivar-group)"
        /mol_type="mRNA"
        /cultivar="Nipponbare"
        /db_xref="taxon:39947"
        /clone="OSJUNE0812"
        /tissue_type="Leaf"
        /dev_stage="3 week"
        /lab_host="DH10B"
        /clone_lib="OSJUNE"
        /note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:
        XhoI; 6 hrs after inoculation with Rice Blast (C9240-1)"
  BASE COUNT    213 a   206 c   176 g   185 t
  ORIGIN
    Query Match      73.3%; Score 493; DB 14; Length 780;
    Best Local Similarity 99.7%; Pred. No. 7.4e-161;
    Matches 593; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
  QY 35 TCTTGGGCTGTGGGTGAGCAGACCAAGAAATTCCTCCCTCAGTGGTTTGAGCTGTCCA 94
  Db 676 TCTTGGGCTGTGGGTGAGCAGACCAAGAAATTCCTCCCTCAGTGGTTTGAGCTGTCCA 617
  QY 95 TCCGCTCTTCCATTCATTGGAATGCTGAGGAAATCTGTCAACATGCCCAAGACTGCCAT 154
  Db 616 TCCGCTCTTCCATTCATTGGAATGCTGAGGAAATCTGTCAACATGCCCAAGACTGCCAT 557
  QY 155 GGCATTACCATACAGCTCCATTGTTGGTCAGACATCGGTGCGAGGCGGAGCGCAT 214
  Db 556 GGCATTACCATACAGCTCCATTGTTGGTCAGACATCGGTGCGAGGCGGAGCGCAT 497
  QY 215 TCGTCTGAAGGACATGGCTGCAAGAGAGCGAGCTGATTTCACACCGTGGTGCATGTA 274
  Db 496 TCGTCTGAAGGACATGGCTGCAAGAGAGCGAGCTGATTTCACACCGTGGTGCATGTA 437
  QY 275 TCCAAACAAGACTGCAAAATTCAGTCAGTCACCGAGGCAAGCATGGGATCCGCTCGCAT 334
  Db 436 TCCAAACAAGACTGCAAAATTCAGTCAGTCACCGAGGCAAGCATGGGATCCGCTCGCAT 377
  QY 335 GAAGATGATGCGGAGAGCGGCTTCTGGTGTGCTGTGCTCCAAACACCAAGCATGTGTTT 394
  Db 376 GAAGATGATGCGGAGAGCGGCTTCTGGTGTGCTGTGCTCCAAACACCAAGCATGTGTTT 317
  QY 395 CTGATTGCTCACTGATTGGAAAAATTTGTATCTACAGTATCCCTGGAGAGTGAGATTG 454
  Db 316 CTGATTGCTCACTGATTGGAAAAATTTGTATCTACAGTATCCCTGGAGAGTGAGATTG 257
  QY 455 ATATTGAGTCTATTTTATCTTGTGATGTAATTTGCTTGTGCTCCCTCAGAACTATTTCG 514

```

Db 256 ATATTGAGTCTATTTTATCTGTGATGTAATTGCTTTGCTTGTCTCCCTCAGAGTATTCG 197
 QY 515 TTGTTGTTGGGATGAGCAAGTGGAAATAGAGTGTCTACTATATACACGATCAITCTGTT 574
 Db 196 TTGTTGTTGGGATGAGCAAGTGGAAATAGAGTGTCTACTATATACACGATCAITCTGTT 137
 QY 575 GTTAAGTTTGGCAGTTCTGAGTTCATGATCTATCTGTAATTTGATGAGTGGATTT 629
 Db 136 GTTAAGTTTGGCAGTTCTGAGTTCATGATCTATCTGTAATTTGATGAGTGGATTT 82

RESULT 6
 CB655085/c
 LOCUS
 DEFINITION OSJNEC08C12.r OSJNEC Oryza sativa (japonica cultivar-group) cDNA
 clone OSJNEC08C12 3', mRNA sequence.
 ACCESSION CB655085
 VERSION CB655085.1 GI:29658810
 KEYWORDS EST.

SOURCE
 ORGANISM Oryza sativa (japonica cultivar-group)
 Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 1 (bases 1 to 727)
 Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
 Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
 Large-scale identification of ESTs involved in the interaction
 between rice and Magnaporthe grisea
 Unpublished
 Contact: Rod Wing
 Arizona Genomics Institute
 University of Arizona
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
 85721-0088, USA
 Tel: 520 626 3967
 Fax: 520 621 9288
 Email: http://genome.arizona.edu
 PCR Primers
 FORWARD: gta aaa cgg cca gtc
 BACKWARD: gga aac agc tat gac cat g
 Plate: 08 row: C column: 12
 Seq primer: gga aac agc tat gac cat g.
 Location/Qualifiers
 1..727

FEATURES
 source
 1..727
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="OSJNEC08C12"
 /tissue_type="Leaf"
 /dev_stage="3 week"
 /lab_host="DH10S"
 /clone_lib="OSJNEC"
 /note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:
 XhoI; 6 hrs after inoculation with Rice Blast (C9240-1)"
 BASE COUNT 204 a 192 c 166 g 165 t

Query Match 73.1%; Score 492; DB 14; Length 727;
 Best Local Similarity 99.7%; Pred. No. 1.7e-160;
 Matches 592; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 35 TCTTGGGTTGGCTGAGCACCACCAAGAAATCTCCCTCAGTGGTTTGAGCTGCCA 94
 Db 618 TCTTGGGTTGGCTGAGCACCACCAAGAAATCTCCCTCAGTGGTTTGAGCTGCCA 559
 QY 95 TGCCCTGTTTCATTCTGGAATCTGAGGAATCTGTCAACATGCCCAAGACTGCCAT 154
 Db 558 TGCCCTGTTTCATTCTGGAATCTGAGGAATCTGTCAACATGCCCAAGACTGCCAT 499
 QY 155 GGCATTACCATAGAGCTCCATTTGTTGGTCAGCAATCGGTCAGGGCGAGCGCAT 214

Db 498 GGCATTACCATAGAGCTCCATTTGTTGGTCAGCAATCGGTCAGGGCGAGCGCAT 439
 QY 215 TCGTCTGAGGCACTGGCTGCAAGAGCGAGCGCTGATTTCCACCACCGTGGCTGACATGTA 274
 Db 438 TCGTCTGAGGCACTGGCTGCAAGAGCGAGCGCTGATTTCCACCACCGTGGCTGACATGTA 379
 QY 275 TCCAAACAAGACTGCAAAATTGCAAGTGCACCGAGGGCAAGCATGGGATCCGCTCGCAT 334
 Db 378 TCCAAACAAGACTGCAAAATTGCAAGTGCACCGAGGGCAAGCATGGGATCCGCTCGCAT 319
 QY 335 GAAGATGATGGGGAGCGGGCTTCTGGTGGTGGTGGTCTCCAAACCAAGCATGTGTTT 394
 Db 318 GAAGATGATGGGGAGCGGGCTTCTGGTGGTGGTGGTCTCCAAACCAAGCATGTGTTT 259
 QY 395 CTGATTGCTCACTGATTGGAAAATTTGATCTACCAAGTATCCCTGGAGAGTGGAGATTG 454
 Db 258 CTGATTGCTCACTGATTGGAAAATTTGATCTACCAAGTATCCCTGGAGAGTGGAGATTG 199
 QY 455 ATATTGAGTCTATTTTATCTTGTGATGTAATTTGCTGTTGTTGCTCCCAAGATATTCG 514
 Db 198 ATATTGAGTCTATTTTATCTTGTGATGTAATTTGCTGTTGTTGCTCCCAAGATATTCG 139
 QY 515 TTGTTGTTGGGATGAGCAAGTGGAAATAGAGTGTCTACTATATACACGATCAITCTGTT 574
 Db 138 TTGTTGTTGGGATGAGCAAGTGGAAATAGAGTGTCTACTATATACACGATCAITCTGTT 79
 QY 575 GTTAAGTTTGGCAGTTCTGAGTTCATGATCTGTAATTTGATGAGTGGATTT 628
 Db 78 GTTAAGTTTGGCAGTTCTGAGTTCATGATCTGTAATTTGATGAGTGGATTT 25

RESULT 7
 CB663774/c
 LOCUS
 DEFINITION OSJNEC09B20.r OSJNEC Oryza sativa (japonica cultivar-group) cDNA
 clone OSJNEC09B20 3', mRNA sequence.
 ACCESSION CB663774
 VERSION CB663774.1 GI:29667499
 KEYWORDS EST.

SOURCE
 ORGANISM Oryza sativa (japonica cultivar-group)
 Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 1 (bases 1 to 763)
 Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
 Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
 Large-scale identification of ESTs involved in the interaction
 between rice and Magnaporthe grisea
 Unpublished
 Contact: Rod Wing
 Arizona Genomics Institute
 University of Arizona
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
 85721-0088, USA
 Tel: 520 626 3967
 Fax: 520 621 9288
 Email: http://genome.arizona.edu
 PCR Primers
 FORWARD: gta aaa cgg cca gtc
 BACKWARD: gga aac agc tat gac cat g
 Plate: 09 row: B column: 20
 Seq primer: gga aac agc tat gac cat g.
 Location/Qualifiers
 1..763

FEATURES
 source
 1..763
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="OSJNEC09B20"
 /tissue_type="Leaf"
 /dev_stage="3 week"
 /lab_host="DH10S"

/clone.lib="OSUNED"
/notes="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
XhoI; 24 hrs after immunization with Rice Blast (C9240-1)"
BASE COUNT 217 a 202 c 177 g 167 t
ORIGIN

Query Match 70.6%; Score 475; DB 14; Length 763;
Best Local Similarity 99.5%; Pred. No. 1.3e-154;
Matches 625; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 35 TCTTGGGTTGGCGTGACACCAAGAAATCTCCCTCAGTGGTTGAGCTGTCCA 94
Db 628 TCTTGGGTTGGCGTGACACCAAGAAATCTCCCTCAGTGGTTGAGCTGTCCA 569
QY 95 TGCCGCTGTTCCATTCATTTGGAAATCTGAGGAAATCTGCAACATGCCAGACTGCCAT 154
Db 568 TGCCGCTGTTCCATTCATTTGGAAATCTGAGGAAATCTGCAACATGCCAGACTGCCAT 509
QY 155 GCATTCACCATAGACGCTCCATTTGGTTCAGCAATCGGGTCGAGGGCGAGCGCAT 214
Db 508 GCATTCACCATAGACGCTCCATTTGGTTCAGCAATCGGGTCGAGGGCGAGCGCAT 449
QY 215 TCGTCTGAAGGCACTGGCTGCAAGAGCGACGCTGATCCACACCGTGGCTGACATGTA 274
Db 448 TCGTCTGAAGGCACTGGCTGCAAGAGCGACGCTGATCCACACCGTGGCTGACATGTA 389
QY 275 TCCRAACAGACTGCAAAATTCAGTGCACACGAGGCGAAGCATGGGATCCGCTGGCAT 334
Db 388 TCCRAACAGACTGCAAAATTCAGTGCACACGAGGCGAAGCATGGGATCCGCTGGCAT 329
QY 335 GAAGATGATGGCGGACGGCTTCTGGTGGTCTGCTGCTCCAAACCAACATGTTGTTT 394
Db 328 GAAGATGATGGCGGACGGCTTCTGGTGGTCTGCTGCTCCAAACCAACATGTTGTTT 269
QY 395 CTGATGCTCACTGATGGAAATTTGATCTACAGATCCCTGGAGTGGAGATTG 454
Db 268 CTGATGCTCACTGATGGAAATTTGATCTACAGATCCCTGGAGTGGAGATTG 209
QY 455 ATATTGACTCTATTTTATCTTGTGATTAATTGCTTGTCTGCTCCCTCAGAGTATTCG 514
Db 208 ATATTGACTCTATTTTATCTTGTGATTAATTGCTTGTCTGCTCCCTCAGAGTATTCG 149
QY 515 TTGTTTGGGATGAGACAGTGGAAATTAAGTGGTCTACTATATACAGATTCATTCGTT 574
Db 148 TTGTTTGGGATGAGACAGTGGAAATTAAGTGGTCTACTATATACAGATTCATTCGTT 89
QY 575 GTTAAGTTTGGCAGTTCTGCAAGTTCATGTAATCTGTAATTTGATGATGCTGATTCCT 634
Db 88 GTTAAGTTTGGCAGTTCTGCAAGTTCATGTAATCTGTAATTTGATGATGCTGATTCCT 29
QY 635 ATTTATCAATCGTCATTTACTGTTGT 662
Db 28 ATTTATCAATCGTCATTTACTGTTGT 1

RESULT 8
AU165747
LOCUS
DEFINITION
AU165747 Rice panicle at flowering stage Oryza sativa (japonica
cultivar-group) cDNA clone B4287, mRNA sequence.
AU165747
VERSION
KEYWORDS
SOURCE
ORGANISM
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 705)
Sasaki, T. and Yamamoto, K.
Rice cDNA from panicle at flowering stage (2000)
Unpublished
JOURNAL
COMMENT
Contact: Takuji Sasaki

National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsasaki@affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/
PROJECT = RGP

FEATURES
source
1..705
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="B4287"
/stage="flowering stage"
/clone_lib="Rice panicle at flowering stage"
/note="Organ: panicle; Rice cDNA from panicle at flowering stage"

BASE COUNT 175 a 150 c 178 g 199 t 3 others
ORIGIN
Query Match 59.4%; Score 467; DB 9; Length 705;
Best Local Similarity 99.6%; Pred. No. 7.9e-152;
Matches 567; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 105 CCATTTCATTGGAAATGCTGAGGAAATCTGTCAACATGCCCAAGACTGCCATGCCATCCACC 164
Db 131 CCATTTCATTGGAAATGCTGAGGAAATCTGTCAACATGCCCAAGACTGCCATGCCATCCACC 190
QY 165 ATAGCAGCTCCATTTGTTGTCAGCAATCTGGTTCAGGGCGGAGCGCATTCGTTCTGAAG 224
Db 191 ATAGCAGCTCCATTTGTTGTCAGCAATCTGGTTCAGGGCGGAGCGCATTCGTTCTGAAG 250
QY 225 GCATCTGGCTGCAAGAGCGAGCGCTGATCCACCAACCGTGGCTGACATGATCCAAACAG 284
Db 251 GCATCTGGCTGCAAGAGCGAGCGCTGATCCACCAACCGTGGCTGACATGATCCAAACAG 310
QY 285 ACTGCCAATTCAGTGCACCGGAGCGGAGCGATGGATCCGCTCCGATGAAGATGATG 344
Db 311 ACTGCCAATTCAGTGCACCGGAGCGGAGCGATGGATCCGCTCCGATGAAGATGATG 370
QY 345 GCGGACGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 404
Db 371 GCGGACGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 430
QY 405 ACTGATGGAAATTTGTTATCTACAGTATCCCTGGAGAGTGGAGATGATATGAGTC 464
Db 431 ACTGATGGAAATTTGTTATCTACAGTATCCCTGGAGAGTGGAGATGATATGAGTC 490
QY 465 TATTTTATCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 524
Db 491 TATTTTATCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 550
QY 525 GATGAGACAGTGGAAATGAGTGTCTACTATATACAGATCACTGTTGTTGTTGTTGTTGTTGTTG 584
Db 551 GATGAGACAGTGGAAATGAGTGTCTACTATATACAGATCACTGTTGTTGTTGTTGTTGTTGTTG 610
QY 585 CCAGTTCTCAGTCTCATGATCTGTAATTTGATGATGATGATGATGATGATGATGATGATGATG 644
Db 611 CCAGTTCTCAGTCTCATGATCTGTAATTTGATGATGATGATGATGATGATGATGATGATGATG 670
QY 645 COTCATTTACTGCTGTGTAATAAAAAA 673
Db 671 CGTCATTACTGCTGTGTAATAAAAAA 699

RESULT 9
AU031690
LOCUS
DEFINITION
AU031690 Rice root Oryza sativa (japonica cultivar-group) cDNA
clone R0565_62, mRNA sequence.
AU031690
ACCESSION
VERSION
AU031690.1 GI:3767663

KEYWORDS
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.
AUTHORS Minobe, Y. and Sasaki, T.
JOURNAL Rice cDNA from root
COMMENT Unpublished
 Contact: Takuji Sasaki
 National Institute of Agrobiological Resources
 Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
 305-8602, Japan 7441
 Fax: 81-298-38-7468
 Email: tsasaki@ab.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/
TITLE PROJECT = "RGP"
FEATURES POLYA=No.
 Location/Qualifiers
 1..718
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="rRNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="R0565.62"
 /clone_lib="Rice root"
 /note="Prepared from seedling root."
BASE COUNT 177 a 158 c 176 g 204 t 3 others
ORIGIN
 Query Match 69.2%; Score 456; DB 9; Length 718;
 Best Local Similarity 99.6%; Pred. No. 1.7e-151;
 Matches 566; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 97 CGCTGTTCCATTCATTGGAATGCTGAGGAATCTGTCAACATGCCCAAGACTGCCATGG 156
 Db 124 CGCTGTTCCATTCATTGGAATGCTGAGGAATCTGTCAACATGCCCAAGACTGCCATGG 183
 QY 157 CATTACATAGCAGCTCCATTGTTGTCAGACATCGGTTCGAGCGCGAGCGCATTC 216
 Db 184 CATTACATAGCAGCTCCATTGTTGTCAGACATCGGTTCGAGCGCGAGCGCATTC 243
 QY 217 GTCTGAAGCCTGCTGCTGCAAGAGCGAGCGGTGATTCACACACCGTGGCTGACATGTATC 276
 Db 244 GTCTGAAGCCTGCTGCTGCAAGAGCGAGCGGTGATTCACACACCGTGGCTGACATGTATC 303
 QY 277 CAACAGAGCTCAATTTGCAGTGACACGAGGCAAGCGATCGGTCCGTCGGATGA 336
 Db 304 CAACAGAGCTCAATTTGCAGTGACACGAGGCAAGCGATCGGTCCGTCGGATGA 363
 QY 337 AGATGATGCGGAGCGGCTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTTCT 396
 Db 364 AGATGATGCGGAGCGGCTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTTCT 423
 QY 397 GATTGCTCACTGATTGGAATTTCTATCTACCTACCTACCTACCTACCTACCTACCTACCT 456
 Db 424 GATTGCTCACTGATTGGAATTTCTATCTACCTACCTACCTACCTACCTACCTACCTACCT 483
 QY 457 ATTGAGTCTATTATTCTTTGATGTAATTTGCTTTGCTTTGCTTTGCTTTGCTTTGCTTTG 516
 Db 484 ATTGAGTCTATTATTCTTTGATGTAATTTGCTTTGCTTTGCTTTGCTTTGCTTTGCTTTG 543
 QY 517 TGTTCGTGGSATGAGACAGTGGATTAAGAGTCTACTATATACACGATCAATTCGTTGT 576
 Db 544 TGTTCGTGGSATGAGACAGTGGATTAAGAGTCTACTATATACACGATCAATTCGTTGT 603
 QY 577 TAAGTTTGGCAGTTCTGCGAGTTCTGATATCTGTAATTTGATGATGCTGATTTCTACTAT 636
 Db 604 TAAGTTTGGCAGTTCTGCGAGTTCTGATATCTGTAATTTGATGATGCTGATTTCTACTAT 663
 QY 637 TTATCAATCGTCAATTAATCTGTGTGTA 664

664 TTATCAATCGTCAATTAATCTGTGTGTA 691
RESULT 10
LOCUS CB619892/c
DEFINITION Oryza sativa (indica cultivar-group) cDNA
ACCESSION CB619892
VERSION CB619892.1
KEYWORDS EST.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 850)
AUTHORS Jantasuriyarat, C., Lu, G., Gowda, M., Hatfield, J., Zhou, B., Mazur, E., Kudrna, D., Dean, R., Soderlund, C., Wang, R. and Wang, G.
TITLE Large-scale identification of ESTs involved in the interaction between rice and Magnaporthe oryzae
JOURNAL Unpublished
COMMENT Contact: Rod Wing
 Arizona Genomics Institute
 University of Arizona
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
 85721-0088, USA
 Tel: 520 626 3967
 Fax: 520 621 9288
 Email: http://genome.arizona.edu
PCR PRIMERS
 FORWARD: gca aaa cga cgg cca gtc
 BACKWARD: gga aac agc tat gac cat g
 Plate: 04 row: M column: 03
 Seq primer: gga aac agc tat gac cat g.
FEATURES Location/Qualifiers
 1..850
 /organism="Oryza sativa (indica cultivar-group)"
 /mol_type="rRNA"
 /cultivar="IR36"
 /db_xref="taxon:39948"
 /clone="OSIIEA04M03"
 /tissue_type="Leaf"
 /dev_stage="3 week"
 /lab_host="DH10B"
 /clone_lib="OSIIEA"
 /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2: XhoI; Lesion Mimic SPL11"
BASE COUNT 218 a 232 c 206 g 194 t
ORIGIN
 Query Match 68.8%; Score 463; DB 14; Length 850;
 Best Local Similarity 99.6%; Pred. No. 1.8e-150;
 Matches 563; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 35 TCTTGGGCTGGCTGAGCAGCAGCAAGAAATTCCTCCCTCAGTGGTTTGAGCTGTCCA 94
 Db 565 TCTTGGGCTGGCTGAGCAGCAGCAGCAAGAAATTCCTCCCTCAGTGGTTTGAGCTGTCCA 506
 QY 95 TCCCGCTGTTCATTTCATTGGAATGCTGAGGAATCTGTCAACATGCCCAAGACTGCCAT 154
 Db 505 TCCCGCTGTTCATTTCATTGGAATGCTGAGGAATCTGTCAACATGCCCAAGACTGCCAT 446
 QY 155 GGCATTCACCATGAGCAGCTCCATTGTTGTCAGACATCGGTTCGAGGCGGAGCGCAT 214
 Db 445 GGCATTCACCATGAGCAGCTCCATTGTTGTCAGACATCGGTTCGAGGCGGAGCGCAT 386
 QY 215 TCGTCTGAGGCACTGGCTGCAAGAGGACGCTGATTCACACCGTGGTGTGACATGTA 274
 Db 385 TCGTCTGAGGCACTGGCTGCAAGAGGACGCTGATTCACACCGTGGTGTGACATGTA 326
 QY 275 TCCAAACAGACTGCAAAATTCAGTGCACCGAGGCGGAGGATCGGCTCGCGAT 334

```

Db      325 TCCAAACAGACTGGAAATTCAGTGCACACGAGGGCAAGGCATGGGATCGCTCGCGAT 266
Qy      335 GAAGATGATGCGGAGCGGGCTTCTGGTGGTGGTCTCTCCAAACACCAAGCATGTGTTT 394
Db      265 GAAGATGATGCGGAGCGGGCTTCTGGTGGTGGTCTCTCCAAACACCAAGCATGTGTTT 206
Qy      395 CTGATTGCTCACTGATTCGAAATTCATCTACAGTATCCCTGGAGAGTGGAGATTG 454
Db      205 CTGATTGCTCACTGATTCGAAATTCATCTACAGTATCCCTGGAGAGTGGAGATTG 146
Qy      455 ATATTGAGTCTATTTATCTTTGATGTAATTCGCTTTGCTTGTCCCTCAGAAGTATTG 514
Db      145 ATATTGAGTCTATTTATCTTTGATGTAATTCGCTTTGCTTGTCCCTCAGAAGTATTG 86
Qy      515 TTGTTTGGGATGACAGAGTGGAAATTCATCTACAGTATCCCTGGAGAGTGGAGATTG 574
Db      85 TTGTTTGGGATGACAGAGTGGAAATTCATCTACAGTATCCCTGGAGAGTGGAGATTG 26
Qy      575 GTTAAGTTTCCAGTTCTGCAAGTTC 599
Db      25 GTTAAGTTTCCAGTTCTGCAAGTTC 1

RESULT 11
CB647260/c
LOCUS      792 bp mRNA linear EST 08-APR-2003
DEFINITION OSJNEB10B21.r OSJNEB Oryza sativa (japonica cultivar-group) cDNA
            clone OSJNEB10B21 3', mRNA sequence.
ACCESSION CB647260
VERSION    CB647260.1 GI:29642253
KEYWORDS   EST.
SOURCE     Oryza sativa (japonica cultivar-group)
ORGANISM   Oryza sativa (japonica cultivar-group)
            Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
            1 (bases 1 to 792)
REFERENCE  Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
            Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
            Large-scale identification of ESTs involved in the interaction
            between rice and Magnaporthe grisea
            Unpublished
            Contact: Rod Wing
            Arizona Genomics Institute
            University of Arizona
            Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
            85721-0088, USA
            Tel: 520 626 3967
            Fax: 520 621 9288
            Email: http://genome.arizona.edu
PCR Primers
FORWARD: gta aaa cga cgg cca gtc
BACKWARD: gga aac agc tat gac cat g
Plate: 10 row: B column: 21
Seq primer: gga aac agc tat gac cat g.
            Location/Qualifiers
FEATURES
    source
        1..792
            /organism="Oryza sativa (japonica cultivar-group)"
            /mol_type="mRNA"
            /cultivar="Nipponbare"
            /db_xref="taxon:39947"
            /clone="OSJNEB10B21"
            /tissue_type="Leaf"
            /dev_stage="3 week"
            /lab_host="DH10B"
            /clone_lib="OSJNEB"
            /notes="Vector: pBluescript II KS +; Site 1: Scori; Site 2:
            XhoI; 24 hrs after inoculation with Rice Blast (Che 86061
            )"
BASE COUNT      222 a      212 c      184 g      174 t
ORIGIN
Query Match      65.8%; Score 443; DB 14; Length 792;

```

```

Best Local Similarity 99.5%; Pred. No. 1.5e-143;
Matches 593; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy      35 TCTTGGGGTGGCGTGCAGCACACCAAGAAATTCCTCCCTCAGTGTGTTTGCAGCTGTCCA 94
Db      629 TCTTGGGGTGGCGTGCAGCACACCAAGAAATTCCTCCCTCAGTGTGTTTGCAGCTGTCCA 570
Qy      95 TCCGCTGTTCCTATTCATTGGAATCTGAGGAAATCTGTCAACATGCCCAAGACTGCCAT 154
Db      569 TCCGCTGTTCCTATTCATTGGAATCTGAGGAAATCTGTCAACATGCCCAAGACTGCCAT 510
Qy      155 GGCATTCACCATAGCAGCTCCATTTGGTTCAGACCAATCGGGTCGAGGCGGAGCGCAT 214
Db      509 GGCATTCACCATAGCAGCTCCATTTGGTTCAGACCAATCGGGTCGAGGCGGAGCGCAT 450
Qy      215 TCGTCTGAAGGCACTGGCTGCAAGAGGACGCTGATTCACCACTGGTGGCTGACATGTA 274
Db      449 TCGTCTGAAGGCACTGGCTGCAAGAGGACGCTGATTCACCACTGGTGGCTGACATGTA 390
Qy      275 TCCAAACAGACTGCATAATTCAGTGACACCGAGGGCAAGGCATGGATCCGCTCGCAT 334
Db      389 TCCAAACAGACTGCATAATTCAGTGACACCGAGGGCAAGGCATGGATCCGCTCGCAT 330
Qy      335 GAAGATGATGCGGAGCGGGCTTCTGGTGGTGGTCTCTCCAAACACCAAGCATGTGTTT 394
Db      329 GAAGATGATGCGGAGCGGGCTTCTGGTGGTGGTCTCTCCAAACACCAAGCATGTGTTT 270
Qy      395 CTGATTGCTCACTGATTCGAAATTCATCTACAGTATCCCTGGAGAGTGGAGATTG 454
Db      269 CTGATTGCTCACTGATTCGAAATTCATCTACAGTATCCCTGGAGAGTGGAGATTG 210
Qy      455 ATATTGAGTCTATTTATCTTTGATGTAATTCGCTTTGCTTGTCCCTCAGAAGTATTG 514
Db      209 ATATTGAGTCTATTTATCTTTGATGTAATTCGCTTTGCTTGTCCCTCAGAAGTATTG 150
Qy      515 TTGTTTGGGATGACAGAGTGGAAATTCATCTACAGTCTACTATATACAGATCAITCTGTT 574
Db      149 TTGTTTGGGATGACAGAGTGGAAATTCATCTACAGTCTACTATATACAGATCAITCTGTT 90
Qy      575 GTTAAGTTTCCAGTTCTGCAAGTTCATCTGATGTAATTCGCTTTGCTTGTCCCTCAGAAGTATTG 630
Db      89 GTTAAGTTTCCAGTTCTGCAAGTTCATCTGATGTAATTCGCTTTGCTTGTCCCTCAGAAGTATTG 34

RESULT 12
CB647623/c
LOCUS      682 bp mRNA linear EST 08-APR-2003
DEFINITION OSJNEB10K01.r OSJNEB Oryza sativa (japonica cultivar-group) cDNA
            clone OSJNEB10K01 3', mRNA sequence.
ACCESSION CB647623
VERSION    CB647623.1 GI:29642616
KEYWORDS   EST.
SOURCE     Oryza sativa (japonica cultivar-group)
ORGANISM   Oryza sativa (japonica cultivar-group)
            Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
            1 (bases 1 to 682)
REFERENCE  Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
            Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
            Large-scale identification of ESTs involved in the interaction
            between rice and Magnaporthe grisea
            Unpublished
            Contact: Rod Wing
            Arizona Genomics Institute
            University of Arizona
            Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
            85721-0088, USA
            Tel: 520 626 3967
            Fax: 520 621 9288
            Email: http://genome.arizona.edu
PCR Primers
FORWARD: gta aaa cga cgg cca gtc
BACKWARD: gga aac agc tat gac cat g
Plate: 10 row: B column: 21
Seq primer: gga aac agc tat gac cat g.
            Location/Qualifiers
FEATURES
    source
        1..682
            /organism="Oryza sativa (japonica cultivar-group)"
            /mol_type="mRNA"
            /cultivar="Nipponbare"
            /db_xref="taxon:39947"
            /clone="OSJNEB10K01"
            /tissue_type="Leaf"
            /dev_stage="3 week"
            /lab_host="DH10B"
            /clone_lib="OSJNEB"
            /notes="Vector: pBluescript II KS +; Site 1: Scori; Site 2:
            XhoI; 24 hrs after inoculation with Rice Blast (Che 86061
            )"
BASE COUNT      222 a      212 c      184 g      174 t
ORIGIN
Query Match      65.8%; Score 443; DB 14; Length 792;

```

BACKWARD: gga aac agc tat gac cat g
 Plate: 10 row: K column: 01
 Seq primer: gga aac agc tat gac cat g
 Location/Qualifiers
 1. 682
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="OSJNEB10K01"
 /tissue_type="Leaf"
 /dev_stage="3 week"
 /lab_host="DH10B"
 /clone_lib="OSJNEB"
 /note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:
 XhoI; 24 hrs after inoculation with Rice Blast (Che 86061)"

FEATURES

source

BASE COUNT 199 a 179 c 145 g 159 t
 ORIGIN

Query Match 65.7%; Score 442; DB 14; Length 682;
 Best Local Similarity 99.5%; Pred. No. 3.6e-143;
 Matches 592; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 35 TCTTGGGTGTGGCTGAGCACCACCAAGAAATTCCTCCCTCAGTGGTTTGACGCTGTCCA 94
 Db 634 TCTTGGGTGTGGCTGAGCACCACCAAGAAATTCCTCCCTCAGTGGTTTGACGCTGTCCA 575
 QY 95 TGCCCTGTTCATTCATGGATCTGAGGAATCTGTCAACATGCTCCACAGATGCCAT 154
 Db 574 TGCCCTGTTCATTCATGGATCTGAGGAATCTGTCAACATGCTCCACAGATGCCAT 515
 QY 155 GGCATTCACATAGCAGCTCCATTTGTTGGTGCAGCAATCGGTCGAGCGGAGCGCAT 214
 Db 514 GGCATTCACATAGCAGCTCCATTTGTTGGTGCAGCAATCGGTCGAGCGGAGCGCAT 455
 QY 215 TGGTCTGAAGCAGCTGGTGTGAAGAGCAGCCTGATTCACACCGTGGTGCATGTA 274
 Db 454 TGGTCTGAAGCAGCTGGTGTGAAGAGCAGCCTGATTCACACCGTGGTGCATGTA 395
 QY 275 TCCAAACAAGACTGCAAAATTCAGTGACACCGAGGCAAGCATGGATCCGTCGGAT 334
 Db 394 TCCAAACAAGACTGCAAAATTCAGTGACACCGAGGCAAGCATGGATCCGTCGGAT 335
 QY 335 GAAGATGATGGCGGACGGGCTTCGTGGTGTGCTGCTCTCAACCAAGCATGTGTT 394
 Db 334 GAAGATGATGGCGGACGGGCTTCGTGGTGTGCTGCTCTCAACCAAGCATGTGTT 275
 QY 395 CTGATTGCTCACTGATTCGAAATTTGATCTACAGTATCCCTGGAGTGAGAGTTG 454
 Db 274 TTGATTGCTCACTGATTCGAAATTTGATCTACAGTATCCCTGGAGTGAGAGTTG 215
 QY 455 ATATTGATGCTATTTATCTTGTGATGAAATTTGCTTGTGCTTGTCCCTCAGAATATCG 514
 Db 214 ATATTGATGCTATTTATCTTGTGATGAAATTTGCTTGTGCTTGTCCCTCAGAATATCG 155
 QY 515 TTTGTTTGTGGATGAGCAAGTGGATTAAGAGTGCTACTATATACAGTATCTTGTGTT 574
 Db 154 TTTGTTTGTGGATGAGCAAGTGGATTAAGAGTGCTACTATATACAGTATCTTGTGTT 95
 QY 575 GTTAAGTTTGCAGTTCTGCAAGTCAATGATCTCTAATTTGATGATGCTGATTT 629
 Db 94 GTTAAGTTTGCAGTTCTGCAAGTCAATGATCTCTAATTTGATGATGCTGATTT 40

RESULT 13

CB627453/c

LOCUS

DEFINITION

OSJNEB02G20.r OSJNEB Oryza sativa (indica cultivar-group) cDNA

clone OSJNEB02G20.3, mRNA sequence.

ACCESSION

CB627453

VERSION

CB627453.1 GI:29622442

KEYWORDS

EST.

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Oryza sativa (indica cultivar-group)
 Oryza sativa (indica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 1 (bases 1 to 777)
 Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
 Kudrna,D., Dean,R., Soderlund,C., Wang,R. and Wang,G.
 Large-scale identification of ESTs involved in the interaction
 between rice and Magnaporthe grisea
 Unpublished
 Contact: Rod Wing
 Arizona Genomics Institute
 University of Arizona
 Biological Sciences West, 488A, P.O. Box 210088, Tucson, AZ
 85721-0088, USA
 Tel: 520 626 3967
 Fax: 520 621 9288
 Email: http://genome.arizona.edu
 PCR Primers
 FORWARD: gta aac cga cgg cca gtc
 BACKWARD: gga aac agc tat gac cat g
 Plate: 02 row: G column: 20
 Seq primer: gga aac agc tat gac cat g.
 Location/Qualifiers

FEATURES

source

1. 777
 /organism="Oryza sativa (indica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="IR36"
 /db_xref="taxon:39946"
 /clone="OSJNEB02G20"
 /tissue_type="Leaf"
 /dev_stage="3 week"
 /lab_host="DH10B"
 /clone_lib="OSJNEB"
 /note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:
 XhoI; 24 hrs after inoculation with Rice Blast (PO6-6-3)"
 BASE COUNT 221 a 206 c 173 g 177 t
 ORIGIN

Query Match 65.7%; Score 442; DB 14; Length 777;
 Best Local Similarity 99.5%; Pred. No. 3.4e-143;
 Matches 592; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 35 TCTTGGGTGTGGCTGAGCACCACCAAGAAATTCCTCCCTCAGTGGTTTGACGCTGTCCA 94
 Db 636 TCTTGGGTGTGGCTGAGCACCACCAAGAAATTCCTCCCTCAGTGGTTTGACGCTGTCCA 577
 QY 95 TGCCCTGTTCATTCATTTGGAATCTGAGGAATCTGTCAACATGCCCAAGACTGCCAT 154
 Db 576 TGCCCTGTTCATTCATTTGGAATCTGAGGAATCTGTCAACATGCCCAAGACTGCCAT 517
 QY 155 GGCATTCACATAGCAGCTCCATTTGTTGGTGCAGCAATCGGTCGAGCGGAGCGCAT 214
 Db 516 GGCATTCACATAGCAGCTCCATTTGTTGGTGCAGCAATCGGTCGAGCGGAGCGCAT 457
 QY 215 TGGTCTGAAGCAGCTGGTGTGAAGAGCAGCCTGATTCACACCGTGGTGCATGTA 274
 Db 456 TGGTCTGAAGCAGCTGGTGTGAAGAGCAGCCTGATTCACACCGTGGTGCATGTA 397
 QY 275 TCCAAACAAGACTGCAAAATTCAGTGACACCGAGGCAAGCATGGATCCGTCGGAT 334
 Db 396 TCCAAACAAGACTGCAAAATTCAGTGACACCGAGGCAAGCATGGATCCGTCGGAT 337
 QY 335 GAAGATGATGGCGGACGGGCTTCGTGGTGTGCTGCTCTCAACCAAGCATGTGTTT 394
 Db 336 GAAGATGATGGCGGACGGGCTTCGTGGTGTGCTGCTCTCAACCAAGCATGTGTTT 277
 QY 395 CTGATTGCTCACTGATTCGAAATTTGATCTACAGTATCCCTGGAGTGAGAGTTG 454
 Db 276 CTGATTGCTCACTGATTCGAAATTTGATCTACAGTATCCCTGGAGTGAGAGTTG 217
 QY 455 ATATTGATGCTATTTATCTTGTGATGTAATTTGCTTGTGCTTGTCCCTCAGAATATTCG 514

```

Db      216 ATATTGAGTCTATTTATCTTGTGATGTAAATGGCTTGTCTGCTCCCTCAGAAGTATTCG 157
QY      515 TTGTTTGTGGGATGAGACAAGTGGAAATAGAGTGTCTACTATPATACAGCATCTCTGTT 574
Db      156 TTGTTTGTGGGATGAGACAAGTGGAAATAGAGTGTCTACTATATACAGCATCTCTGTT 97
QY      575 GTTAAGTTTGGCCAGTTCTGCAGTTCATGTCTGTAAATTTGATGTGCTGATTT 629
Db      96 GTTAAGTTTGGCCAGTTCTGCAGTTCATGTCTGTAAATTTGATGTGCTGATTT 42

RESULT 14
CB620522/c
LOCUS
DEFINITION
  CB620522
  Oryza sativa (indica cultivar-group) cDNA
  clone OS11Ba05W20 3', mRNA sequence.
ACCESSION
  CB620522
VERSION
  1
KEYWORDS
  EST.
ORGANISM
  Oryza sativa (indica cultivar-group)
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
  Ehrhartoideae; Oryzeae; Oryza.
  1 (bases 1 to 799)
  Jantasuiyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
  Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
  Large-scale identification of ESTs involved in the interaction
  between rice and Magnaporthe grisea
  Unpublished
  Contact: Rod Wing
  Arizona Genomics Institute
  University of Arizona
  Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
  85721-0088, USA
  Tel: 520 626 3367
  Fax: 520 621 9288
  Email: http://genome.arizona.edu
PCR Primers
  FORWARD: gta aac cga cgg cca gtc
  BACKWARD: gga aac agc tat gac cat g
  Plate: 05 row: M column: 20
  Seq primer: gga aac agc tat gac cat g.
  Location/Qualifiers
  1..799
  /organism="Oryza sativa (indica cultivar-group)"
  /mol_type="mRNA"
  /cultivar="IR36"
  /db_xref="taxon:39946"
  /clone="OS11Ba05W20"
  /tissue_type="Leaf"
  /dev_stage="3 week"
  /lab_host="DH10B"
  /notes="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
  XhoI; Lesion Mimic SPL 11"
  XhoI; 220 c 194 g 178 t

BASE COUNT
ORIGIN
  207 a 220 c 194 g 178 t

Query Match
Best Local Similarity 65.5%; Score 441; DB 14; Length 799;
Matches 491; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      35 TCTTGGGGTGTGGCGTGAGCACACCAAGAAATTCCTCCCTCAGTGGTTTGCAGCTGTCCA 94
Db      549 TCTTGGGGTGTGGCGTGAGCACACCAAGAAATTCCTCCCTCAGTGGTTTGCAGCTGTCCA 490
QY      95 TGGCGCTGTTCCATTCATTGGATGCTGAGGAATCTGTCAACATGCCCAAGACTGCCAT 154
Db      489 TGGCGCTGTTCCATTCATTGGATGCTGAGGAATCTGTCAACATGCCCAAGACTGCCAT 430
QY      155 GGCATTCCACATAGACGCTCCATTGTTGTCAGACAAATCGGTCGAGGCGGCGGCGCAT 214

```

```

Db      429 GGCAITTCACATAGCAGCTCTCATTTGTTGTCAGACAATCGGTCGAGGGCGAGCGCAT 370
QY      215 TCCTCTGAAGACCACTGGCTGCAAGAGCGACGCTGATTCACACCGCTGGCTGACATGTA 274
Db      369 TCGTCTGAAGGCACTGGCTGCAAGAGCGACGCTGATTCACACCGCTGGCTGACATGTA 310
QY      275 TCCAAACACAGCTGCAAAATTTGCAGTGACACCGAGGCGAAGCATGGATCCGCTCGCGAT 334
Db      309 TCCAAACACAGCTGCAAAATTTGCAGTGACACCGAGGCGAAGCATGGATCCGCTCGCGAT 250
QY      335 GAAGATGATGGCGGCGACGGCTTCTGGTGGTCTCTGCTCTCCAAACCAAGCATGTGTTT 394
Db      249 GAAGATGATGGCGGCGACGGCTTCTGGTGGTCTCTGCTCTCCAAACCAAGCATGTGTTT 190
QY      395 CTGATTGCTCACTGATTGGAAAAATTTGATCTACACAGTATCCCTCGAGAGTGGAGATTG 454
Db      189 CTGATTGCTCACTGATTGGAAAAATTTGATCTACACAGTATCCCTCGAGAGTGGAGATTG 130
QY      455 ATATTGAGTCTATTTATCTTGTGATGTAATTCCTTGTGCTTGTCCCTCAGAAGTATTCG 514
Db      129 ATATTGAGTCTATTTATCTTGTGATGTAATTCCTTGTGCTTGTCCCTCAGAAGTATTCG 70
QY      515 TTGTTTGTGGG 526
Db      69 TTGTTTGTGGG 58

RESULT 15
CB659692/c
LOCUS
DEFINITION
  CB659692
  Oryza sativa (japonica cultivar-group)
  clone OSJNEc16O06 3', mRNA sequence.
ACCESSION
  CB659692
VERSION
  1
KEYWORDS
  EST.
ORGANISM
  Oryza sativa (japonica cultivar-group)
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
  Ehrhartoideae; Oryzeae; Oryza.
  1 (bases 1 to 836)
  Jantasuiyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
  Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
  Large-scale identification of ESTs involved in the interaction
  between rice and Magnaporthe grisea
  Unpublished
  Contact: Rod Wing
  Arizona Genomics Institute
  University of Arizona
  Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
  85721-0088, USA
  Tel: 520 626 3367
  Fax: 520 621 9288
  Email: http://genome.arizona.edu
PCR Primers
  FORWARD: gta aac cga cgg cca gtc
  BACKWARD: gga aac agc tat gac cat g
  Plate: 16 row: O column: 06
  Seq primer: gga aac agc tat gac cat g.
  Location/Qualifiers
  1..836
  /organism="Oryza sativa (japonica cultivar-group)"
  /mol_type="mRNA"
  /cultivar="Nipponbare"
  /db_xref="taxon:39947"
  /clone="OSJNEc16O06"
  /tissue_type="Leaf"
  /dev_stage="3 week"
  /lab_host="DH10B"
  /notes="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
  XhoI; 6 hrs after inoculation with Rice Blast (C9240-1)"
  XhoI; 6 hrs after inoculation with Rice Blast (C9240-1)"

```

BASE COUNT		208 a	227 c	201 g	200 t
ORIGIN					
Query Match		65.5%; Score 441; DB 14; Length 836;			
Best Local Similarity		99.8%; Pred. No. 7.3e-143;			
Matches 491; Conservative		0; Mismatches 1; Indels 0; Gaps 0;			
Qy	35	TCTTGGGTGGCGTGGAGCACCAAGAAATTCCTCCCTCAGTGGTTTCAGCTGTCCA			
Db	549	TCTTGGGTGGCGTGGAGCACCAAGAAATTCCTCCCTCAGTGGTTTCAGCTGTCCA			
Qy	95	TGCCTGCTTCATTCATTTGAATGCTGAGGAAATCTGTCAACATGCCCAAGACTGCCAT			
Db	489	TGCCTGCTTCATTCATTTGAATGCTGAGGAAATCTGTCAACATGCCCAAGACTGCCAT			
Qy	155	GSCATTACCATAGCAGCTCCATTGTTGGTCAGACATCGGTGAGGGCGGAGCGCAT			
Db	429	GSCATTACCATAGCAGCTCCATTGTTGGTCAGACATCGGTGAGGGCGGAGCGCAT			
Qy	215	TCGTCTGAAGCACTGGCTGCAAGAGCGACGCTGATTCACACCCTGGCTGACATGTA			
Db	369	TCGTCTGAAGCACTGGCTGCAAGAGCGACGCTGATTCACACCCTGGCTGACATGTA			
Qy	275	TCCAAACAGACTGCNATTCAGTGCACACGAGGGCAAGCATGGATCCGCTCGCGAT			
Db	309	TCCAAACAGACTGCNATTCAGTGCACACGAGGGCAAGCATGGATCCGCTCGCGAT			
Qy	335	GAAGATGATGCGGACCGGCTTCTGGTGTGCTGCTGCTCCACACCAAGCATGTGTT			
Db	249	GAAGATGATGCGGACCGGCTTCTGGTGTGCTGCTGCTCCACACCAAGCATGTGTT			
Qy	395	CTGATTGCTCACTGATTGGAAATTTGTATCTACAGTATCCCTGGAGAGTGGAGATTG			
Db	189	CTGATTGCTCACTGATTGGAAATTTGTATCTACAGTATCCCTGGAGAGTGGAGATTG			
Qy	455	ATATTGATCTATTTTATCTTGTGATGTAATTGCTTTGCTTTGCTTCAGAAATTCG			
Db	129	ATATTGATCTATTTTATCTTGTGATGTAATTGCTTTGCTTTGCTTCAGAAATTCG			
Qy	515	TTTCTTTTGTGG 526			
Db	69	TTTCTTTTGTGG 58			

Search completed: November 14, 2003, 11:33:51
Job time : 1931 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2003, 10:20:28 ; Search time 58 Seconds

(without alignments)
5121.566 Million cell updates/sec

Title: US-09-868-025-1

Perfect score: 673

Sequence: 1 tttactctgctctcgat.....actgtgtgtaaaaaaaaaa 673

Scoring table: OLIGO-NUC

Gapop 60.0 , Gapext 60.0

Searched: 569978 seqs, 220691566 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA.*

- 1: /cgn2_6/ptodata/2/ina/5A-COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B-COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A-COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B-COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PTUS-COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19	2.8	4403765	3	US-09-103-840A-2
2	19	2.8	4411529	3	US-09-103-840A-1
3	18	2.7	974	4	US-08-894-251A-1
4	17	2.5	957	3	US-09-248-335-59
5	17	2.5	1046	3	US-09-368-140A-5
6	17	2.5	1416	3	US-08-866-928B-2
7	17	2.5	1416	4	US-09-685-836-2
8	17	2.5	1701	4	US-09-411-628-5
9	17	2.5	1702	1	US-07-616-022C-1
10	17	2.5	2167	3	US-08-884-324-7
11	17	2.5	2221	4	US-09-301-978C-1
12	17	2.5	2274	4	US-09-220-132-188
13	17	2.5	3080	4	US-09-411-628-3
14	17	2.5	3615	4	US-09-221-017B-971
15	17	2.5	3651	2	US-08-790-374-1
16	17	2.5	4342	1	US-08-436-044-1
17	17	2.5	4342	2	US-08-436-054-1
18	17	2.5	4342	5	PCT-US95-08812-1
19	17	2.5	5852	4	US-09-853-768-10
20	17	2.5	7037	4	US-09-853-768-3
21	17	2.5	18596	3	US-09-318-448-11
22	17	2.5	18597	4	US-09-362-665-8
23	17	2.5	28994	3	US-08-884-324-14
24	16	2.4	260	1	US-08-594-031-45
25	16	2.4	260	1	US-08-594-031-137
26	16	2.4	260	1	US-08-594-031-144
27	16	2.4	263	2	US-08-318-837-3

28 2.4 266 4 US-09-313-294A-703 Sequence 703, Appl
29 2.4 340 3 US-08-836-075A-73 Sequence 73, Appl
30 2.4 390 4 US-09-702-705-1571 Sequence 1571, Ap
31 2.4 390 4 US-09-736-457-1571 Sequence 1571, Ap
32 2.4 510 4 US-09-252-991A-12747 Sequence 12747, A
33 2.4 573 1 US-08-319-590-24 Sequence 24, Appl
34 2.4 573 1 US-08-487-001A-24 Sequence 24, Appl
35 2.4 573 2 US-08-630-822A-24 Sequence 24, Appl
36 2.4 573 2 US-08-711-905-24 Sequence 24, Appl
37 2.4 573 2 US-09-005-069-24 Sequence 24, Appl
38 2.4 591 1 US-08-487-001A-34 Sequence 34, Appl
39 2.4 591 2 US-08-630-822A-34 Sequence 34, Appl
40 2.4 591 2 US-09-005-069-34 Sequence 34, Appl
41 2.4 604 4 US-09-370-838-175 Sequence 175, Appl
42 2.4 771 4 US-09-252-991A-13201 Sequence 13201, A
43 2.4 858 4 US-09-328-352-2870 Sequence 2870, Ap
44 2.4 885 1 US-08-365-103B-3 Sequence 3, Appl
45 2.4 924 1 US-08-365-103B-5 Sequence 5, Appl

ALIGNMENTS

RESULT 1

US-09-103-840A-2

; Sequence 2, Application US/09103840A

; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; APPLICANT: WHITE, Owen R.

; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; TITLE OF INVENTION: TUBERCULOSIS

; FILE REFERENCE: 24366-20007.00

; CURRENT APPLICATION NUMBER: US/09/103,840A

; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 4403765

; TYPE: DNA

; ORGANISM: Mycobacterium tuberculosis

; FEATURE:

; OTHER INFORMATION: CDC 1551

; OTHER INFORMATION: "n" bases at various positions throughout the sequence

; OTHER INFORMATION: represent a, t, c or g

US-09-103-840A-2

Query Match

Best Local Similarity 2.8%; Score 19; DB 3; Length 4403765;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

47 GCGTGCACACACCAAGAA 65

Db 753861 GCGTGCACACCAAGAA 753879

RESULT 2

US-09-103-840A-1

; Sequence 1, Application US/09103840A

; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; APPLICANT: WHITE, Owen R.

; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; TITLE OF INVENTION: TUBERCULOSIS

; FILE REFERENCE: 24366-20007.00

; CURRENT APPLICATION NUMBER: US/09/103,840A

; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37RV
US-09-103-840A-1

Query Match      2.8%; Score 19; DB 3; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 GCGTGAGCACCAAGAAA 65
    |||||
Db 751901 GCGTGAGCACCAAGAAA 751919

RESULT 3
US-08-894-251A-1
; Sequence 1, Application US/08894251A
; Patent No. 6455305
; GENERAL INFORMATION:
; APPLICANT: Shlomo Melmed (Inventor)
; APPLICANT: Lin Pei (Inventor)
; TITLE OF INVENTION: Pituitary-Tumor-Transforming Genes, and
; TITLE OF INVENTION: Related Products
; FILE REFERENCE: P07 39556 (80392)
; CURRENT APPLICATION NUMBER: US/08/894,251A
; PRIOR FILING DATE: 1999-07-23
; PRIOR APPLICATION NUMBER: PCT/US97/21463
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: US 60/031,338
; PRIOR FILING DATE: 1996-11-21
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 974
; TYPE: DNA
; ORGANISM: Rattus rattus
US-08-894-251A-1

Query Match      2.7%; Score 18; DB 4; Length 974;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 656 TGTGTGTAATAAAAAA 673
    |||||
Db 940 TGTGTGTAATAAAAAA 957

RESULT 4
US-09-248-335-59
; Sequence 59, Application US/09248335
; Patent No. 6096504
; GENERAL INFORMATION:
; APPLICANT: MCGONIGLE, BRIAN
; APPLICANT: O'KEEF, DANIEL
; TITLE OF INVENTION: PLANT GLUTATHIONE-S-TRANSFERASE ENZYMES
; FILE REFERENCE: CL-1128-A
; CURRENT APPLICATION NUMBER: US/09/248,335
; CURRENT FILING DATE: 1999-02-10
; EARLIER APPLICATION NUMBER: 08/324,759
; EARLIER FILING DATE: 1997-September-05
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 59
; LENGTH: 967
; TYPE: DNA
; ORGANISM: maize
US-09-248-335-59

Query Match      2.5%; Score 17; DB 3; Length 967;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 657 GTGTGTAATAAAAAA 673
    |||||
Db 1013 GTGTGTAATAAAAAA 1029

RESULT 5
US-09-068-140A-5
; Sequence 5, Application US/09068140A
; Patent No. 6281409
; GENERAL INFORMATION:
; APPLICANT: Mary Rose Woodhead, Mark Andrew Taylor
; APPLICANT: and Rex Michael Brennan
; TITLE OF INVENTION: Blackcurrant Promoters and Genes
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/068,140A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP96/04807
; FILING DATE: No. 6281409ember 4, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Dinner, Dara L.
; REGISTRATION NUMBER: 33,680
; REFERENCE/DOCKET NUMBER: C70237
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5017
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1046 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Ribes nigrum
; STRAIN: Ben Alder
US-09-068-140A-5

Query Match      2.5%; Score 17; DB 3; Length 1046;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 657 GTGTGTAATAAAAAA 673
    |||||
Db 1013 GTGTGTAATAAAAAA 1029

RESULT 6
US-08-866-928B-2/c
; Sequence 2, Application US/08866928B
; Patent No. 6159702
; GENERAL INFORMATION:
; APPLICANT: Traish, Abdulmaged M
; TITLE OF INVENTION: In-vivo diagnostic methods for determining
```

; TITLE OF INVENTION: whether a primary breast tumor is clinically metastatic
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David Prashker, P.C.
; STREET: P.O. Box 5387
; CITY: Magnolia
; STATE: MA
; COUNTRY: USA
; ZIP: 01930
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.40 Mb storage
; COMPUTER: IBM PS1
; OPERATING SYSTEM: MS DOS
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/866,9288
; FILING DATE: May 30, 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 509,570
; FILING DATE: July 31, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: David Prashker
; REGISTRATION NUMBER: 29,693
; REFERENCE/DOCKET NUMBER: BUMC-060
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (978) 525-3794
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1416
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-866-928B-2
;
; Query Match 2.5%; Score 17; DB 3; Length 1416;
; Best Local Similarity 100.0%; Pred. No. 52;
; Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; Qy 358 CTGGTGGTGGCTGCTGCT 374
; Db 90 CTGGTGGTGGCTGCTGCT 74
;
; RESULT 7
; US-09-685-836-2/c
; Sequence 2, Application US/09685836
; Patent No. 6489113
; GENERAL INFORMATION:
; APPLICANT: Traish, Abdulmaged M
; TITLE OF INVENTION: In-vivo diagnostic methods for determining whether a primary breast tumor is clinically metastatic
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David Prashker, P.C.
; STREET: P.O. Box 5387
; CITY: Magnolia
; STATE: MA
; COUNTRY: USA
; ZIP: 01930
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.40 Mb storage
; COMPUTER: IBM PS1
; OPERATING SYSTEM: MS DOS
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/685,836
; FILING DATE: 07-Oct-2000
; CLASSIFICATION: Unknown
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 509,570
; FILING DATE: July 31, 1995
; ATTORNEY/AGENT INFORMATION:

; NAME: David Prashker
; REGISTRATION NUMBER: 29,693
; REFERENCE/DOCKET NUMBER: BUMC-060
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (978) 525-3794
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1416
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-685-836-2
;
; Query Match 2.5%; Score 17; DB 4; Length 1416;
; Best Local Similarity 100.0%; Pred. No. 52;
; Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; Qy 358 CTGGTGGTGGCTGCTGCT 374
; Db 90 CTGGTGGTGGCTGCTGCT 74
;
; RESULT 8
; US-09-411-628-5
; Sequence 5, Application US/09411628
; Patent No. 6428994
; GENERAL INFORMATION:
; APPLICANT: University of Southern California
; TITLE OF INVENTION: CDNA, GENOMIC, AND PREDICTED PROTEIN
; TITLE OF INVENTION: SEQUENCES OF LEARNING-INDUCED KINASES
; FILE REFERENCE: 13761-707
; CURRENT APPLICATION NUMBER: US/09/411,628
; CURRENT FILING DATE: 1999-10-01
; EARLIER APPLICATION NUMBER: US 60/102,906
; EARLIER FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1701
; TYPE: DNA
; ORGANISM: Orcytolagus cuniculus
; US-09-411-628-5
;
; Query Match 2.5%; Score 17; DB 4; Length 1701;
; Best Local Similarity 100.0%; Pred. No. 52;
; Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; Qy 657 GTGTGTAATAAAAAAAAAA 673
; Db 205 GTGTGTAATAAAAAAAAAA 221
;
; RESULT 9
; US-07-616-022C-1
; Sequence 1, Application US/07616022C
; Patent No. 5427785
; GENERAL INFORMATION:
; APPLICANT: Ronson, Clive W.
; APPLICANT: Kwiatkowski, Robert W.
; TITLE OF INVENTION: IMPROVING RHIZOSPHERIC
; TITLE OF INVENTION: BACTERIA
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kinney & Lange, P.A.
; STREET: 625 Fourth Avenue South, Suite 1500
; CITY: Minneapolis
; STATE: Minnesota
; COUNTRY: U.S.A.
; ZIP: 55415-1659
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX

OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/616.022C
FILING DATE: 19901121
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Sawicki, Z. Peter
REGISTRATION NUMBER: 30,214
REFERENCE/DOCKET NUMBER: LL23.12-0005
TELEPHONE: (612)339-1863
TELEFAX: (612)339-6580
TELEPHONE: 9103805042
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1702
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
US-07-616-022C-1

Query Match 2.5%; Score 17; DB 1; Length 1702;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 GCCATGGCATTCCACAT 166
DB 930 GCCATGGCATTCCACAT 946

RESULT 10
US-08-884-324-7
Sequence 7, Application US/08884324
Patent No. 6060283
GENERAL INFORMATION:
APPLICANT: Takamori OKURA
APPLICANT: Kakui TORIOE
APPLICANT: Masaki KURIMOTO
TITLE OF INVENTION: GENOMIC DNA ENCODING A POLYPEPTIDE CAPABLE
OF INDUCING THE PRODUCTION OF INTERFERON-
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESS: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/884,324
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 185,305/96
FILING DATE: 27-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, ROGER L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: OKURA=1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:

LENGTH: 2167 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: human
TISSUE TYPE: placenta
FEATURE:
NAME/KEY: exon + 3'UTR
LOCATION: 1..2167
IDENTIFICATION METHOD: B
US-08-884-324-7

Query Match 2.5%; Score 17; DB 3; Length 2167;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 403 TCACTGATTGGAATTT 419
DB 1920 TCACTGATTGGAATTT 1936

RESULT 11
US-09-301-978C-1/c
Sequence 1, Application US/09301978C
Patent No. 6392015
GENERAL INFORMATION:
APPLICANT: Paganiban, Antonito
APPLICANT: Callahan, Mark A.
TITLE OF INVENTION: Method of Identifying Modulators of HIV-1 VPU and GAG
TITLE OF INVENTION: Interaction with U Binding Protein (UBP)
FILE REFERENCE: 960296.95335
CURRENT APPLICATION NUMBER: US/09/301,978C
CURRENT FILING DATE: 1999-04-29
PRIOR APPLICATION NUMBER: 60/083,567
PRIOR FILING DATE: 1998-04-30
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 2221
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: unsure
LOCATION: (1514)
OTHER INFORMATION: n = any nucleotide.
NAME/KEY: unsure
LOCATION: (2066)
OTHER INFORMATION: n = any nucleotide.
US-09-301-978C-1

Query Match 2.5%; Score 17; DB 4; Length 2221;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 TCTGTCAACATGCCCAA 145
DB 1648 TCTGTCAACATGCCCAA 1632

RESULT 12
US-09-220-132-188
Sequence 188, Application US/09220132
Patent No. 6506607
GENERAL INFORMATION:
APPLICANT: Sdylian, Andrew W.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT
OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CANC
FILE REFERENCE: 07334-074001
CURRENT APPLICATION NUMBER: US/09/220,132
CURRENT FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: US 60/079,303

```

; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: US 60/069,821
; PRIOR FILING DATE: 1997-12-24
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 188
; LENGTH: 2274
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-220-132-188

Query Match      2.5%; Score 17; DB 4; Length 2274;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 TCTGTCACATGCCCAA 145
Db 595 TCTGTCACATGCCCAA 611

RESULT 13
US-09-411-628-3
; Sequence 3, Application US/09411628
; Patent No. 6428994
; GENERAL INFORMATION:
; APPLICANT: University of Southern California
; TITLE OF INVENTION: CDNA, GENOMIC, AND PREDICTED PROTEIN
; TITLE OF INVENTION: SEQUENCES OF LEARNING-INDUCED KINASES
; FILE REFERENCE: 13761-707
; CURRENT APPLICATION NUMBER: US/09/411,628
; CURRENT FILING DATE: 1999-10-01
; EARLIER APPLICATION NUMBER: US 60/102,906
; EARLIER FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 3080
; TYPE: DNA
; ORGANISM: Oryctolagus cuniculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (249)...(1949)
US-09-411-628-3

Query Match      2.5%; Score 17; DB 4; Length 3080;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 657 GTGTGTAATAAAAAA 673
Db 453 GTGTGTAATAAAAAA 469

RESULT 14
US-09-221-017B-971/c
; Sequence 971, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: ROBB, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: IBM Compatible
; SOFTWARE: FastSeq for Windows Version 2.0b

; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: US/09/221,017B
; FILING DATE: 23-DEC-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1182
; FILING DATE: 31-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1546
; FILING DATE: 30-JAN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP2911
; FILING DATE: 09-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU98/01023
; FILING DATE: 10-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Monicoy, Gladys H
; REGISTRATION NUMBER: 32,430
; REFERENCE/DOCKET NUMBER: 27340-20021.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-813-5600
; TELEFAX: 650-494-0792
; TELEX: 708141
; INFORMATION FOR SEQ ID NO: 971:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3615 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: UNKNOWN
; ORIGINAL SOURCE:
; ORGANISM: PORPHYROMONAS GINGIVALIS
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1...3615
US-09-221-017B-971

Query Match      2.5%; Score 17; DB 4; Length 3615;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 509 TATTCGTTTGTGG 525
Db 1207 TATTCGTTTGTGG 1191

RESULT 15
US-08-790-374-1
; Sequence 1, Application US/08790374
; Patent No. 5863734
; GENERAL INFORMATION:
; APPLICANT: Karayiorgou, Maria
; APPLICANT: Gogos, Joseph A.
; TITLE OF INVENTION: METHODS OF TREATMENT FOR OBSESSIVE-COMPULSIVE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/790,374
; FILING DATE:
```

```

/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Jackson Esq., David A.
/ REGISTRATION NUMBER: 26,742
/ REFERENCE/DOCKET NUMBER: 600-1-189
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 201-487-5800
/ TELEFAX: 201-343-1684
/ TELEX: 133521
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 3651 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ HYPOTHETICAL: NO
/ US-08-790-374-1

Query Match      2.5%; Score 17; DB 2; Length 3651;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      358 CTGGTGGTGGTGGTGGT 374
      |||||
DB      1395 CTGGTGGTGGTGGTGGT 1411

```

Search completed: November 14, 2003, 11:35:21
Job time : 71 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2003, 09:20:07 ; Search time 221 Seconds
(without alignments)
8220.461 Million cell updates/sec

Title: US-09-868-025-1
Perfect score: 673
Sequence: 1 ttatcttgcctgcctgat.....actgtgtataaaaaaaa 673

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2552756 seqs, 1349719017 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : 1_ /SIDS1/gcgdata/geneq/geneq-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneq/geneq-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneq/geneq-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneq/geneq-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneq/geneq-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneq/geneq-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneq/geneq-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneq/geneq-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneq/geneq-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneq/geneq-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneq/geneq-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneq/geneq-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneq/geneq-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneq/geneq-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneq/geneq-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneq/geneq-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneq/geneq-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneq/geneq-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneq/geneq-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneq/geneq-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneq/geneq-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneq/geneq-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneq/geneq-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneq/geneq-emb1/NA2002.DAT.*
25: /SIDS1/gcgdata/geneq/geneq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	673	100.0	673	AAH19867	Rice AGT-SAL 11 po
2	20	3.0	503	ABQ28226	Oligonucleotide fo
3	20	3.0	503	ABQ28227	Oligonucleotide fo
4	20	3.0	569	ABQ53592	Oligonucleotide fo
5	20	3.0	569	ABQ53593	Oligonucleotide fo
6	20	3.0	1967	AAS22945	DNA encoding novel
7	19	2.8	525	AAFO7706	Fusarium venenatum
8	19	2.8	720	AB876857	Frog embryonic gen

9	19	2.8	720	24	AB876869	Frog embryonic gen
10	19	2.8	2057	22	AAF72803	Secreted protein g
11	19	2.8	3974	23	ABL21565	Drosophila melanog
12	19	2.8	7537	22	AAS36235	Human cardiovascular
13	19	2.8	7537	22	AAK65643	Human immune/haema
14	19	2.8	15366	23	ABL21564	Drosophila melanog
15	19	2.8	156638	24	AQ81850	Bifidobacterium lo
16	19	2.8	349980	24	AQ81850	Bifidobacterium lo
17	19	2.8	4403765	22	AA199683	Mycobacterium tube
18	19	2.8	4411529	22	AA199682	Mycobacterium tube
19	18	2.7	178	24	ABK89158	Expressed sequence
20	18	2.7	209	24	ABQ94838	Tumour suppression
21	18	2.7	287	24	ABL72780	Corn tassal-derive
22	18	2.7	342	24	ABN24232	Human ORFX polynuc
23	18	2.7	387	25	ABX48169	Bovine EST associa
24	18	2.7	556	25	ABX57116	Arabidopsis thalia
25	18	2.7	559	22	AAH88746	Pyruvate dehydroge
26	18	2.7	569	25	ABX56960	Arabidopsis thalia
27	18	2.7	597	23	ABL28553	Drosophila melanog
28	18	2.7	599	22	AAK88744	Human digestive sy
29	18	2.7	599	22	AAS31778	Human liver associ
30	18	2.7	599	24	ABN90133	Human liver antige
31	18	2.7	610	24	ABQ98970	Human ORF777 codin
32	18	2.7	649	24	ABQ57106	Human colon cancer
33	18	2.7	764	21	AAK08349	Mou8e TGC-440 8scr
34	18	2.7	792	22	ABA21285	Human nervous syst
35	18	2.7	974	19	AAV36963	Rat pituitary-tumo
36	18	2.7	974	24	ABN87388	Rat pituitary tumo
37	18	2.7	974	24	ABN30908	Rat pituitary CDNA. R
38	18	2.7	974	24	ABN30908	Rat pituitary tumo
39	18	2.7	974	24	ABN30908	Rat pituitary tumo
40	18	2.7	975	21	ABF13716	Aspergillus oryzae
41	18	2.7	1066	22	AAS31201	Human CDNA encodin
42	18	2.7	1066	24	ABQ66525	Human polynucleoti
43	18	2.7	1270	22	ABA21504	Human nervous syst
44	18	2.7	1287	23	ABL10643	Drosophila melanog
45	18	2.7	1311	20	AAK61447	DNA encoding a hum

ALIGNMENTS

RESULT 1
AAH19867
ID AAH19867 standard; DNA; 673 BP.
XX AAH19867;
AC AAH19867;
XX 03-AUG-2001 (first entry)
XX Rice AGT-SAL 11 polynucleotide sequence SEQ ID NO:1.
DE Rice AGT-SAL 11; salt tolerance; plant; cytostatic; antiHIV;
XX proteinase inhibitor; cancer; human immunodeficiency virus;
KW HIV infection; animal disorder; food processing; enzyme industry;
KW biological preservative; ds.
XX Oryza sativa.
OS Oryza sativa.
FN WO200130990-A2.
XX 03-MAY-2001.
XX 11-OCT-2000; 2000MO-IN00099.
XX 13-OCT-1999; 99IN-0000997.
XX (AVES-) AVESTHAGEN GRAINE TECHNOLOGIES PVT LTD.
XX Patell VM, Antony CM, Chandran D, Madurappa A;
XX WPI; 2001-308632/32.
XX P-PSDB, AAB75128.

XX Nucleotide sequence encoding an AGT-SAL 11 polypeptide similar to
PT Bowman Birk II type proteinase inhibitors is useful to confer salt
PT resistance to plants -
XX
XX Claim 1, Page 17; 22pp; English.
XX
XX The present sequence encodes the rice AGT-SAL 11 protein. The AGT-SAL 11
CC protein can be used to confer salt tolerance to plants and other
CC organisms. The AGT-SAL 11 gene was isolated from salt-stressed rice.
CC Also described are: (i) a transgenic plant comprising a recombinant
CC expression cassette comprising a plant promoter operably linked to N1;
CC and (ii) conferring salt tolerance on a plant, comprising introducing
CC the above expression cassette. The AGT-SAL 11 protein is a proteinase
CC inhibitor. The proteinase inhibitor may be used to confer stress
CC tolerance to many plants and organisms including cotton, maize, rice,
CC soybean, sugar beet, wheat, fruit, vegetables and vines, particularly
CC biotic bacterial, fungal and pest stresses. Proteinase inhibitors are
CC also useful in the treatment of cancer, human immunodeficiency virus
CC (HIV) infection and other animal disorders. The gene may also be useful
CC in food processing and enzyme industries as an inhibitor of proteinase
CC activity and as a biological preservative.
XX
XX Sequence 673 BP; 163 A; 145 C; 170 G; 195 T; 0 other;
XX SQ
Query Match 100.0%; Score 673; DB 22; Length 673;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 673; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTTACCTTCTCTCTGGATGGGAGCAAACTCCATCTTGGGTGGGTGAGCACCA 60
Db 1 TTTACCTTCTCTCTGGATGGGAGCAAACTCCATCTTGGGTGGGTGAGCACCA 60
QY 61 AGAAATTTCTCCCTCCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 120
Db 61 AGAAATTTCTCCCTCCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 120
QY 121 TGAGGAATCTGTCACATGCCAGCTGCCATGCCATGCCATGCCATGCCATGCCAT 180
Db 121 TGAGGAATCTGTCACATGCCAGCTGCCATGCCATGCCATGCCATGCCATGCCAT 180
QY 181 TTGGTTCAGACAAATCGGTCGAGGCGGAGCGCATTCGTCTGAAGCGCTGGTCAAGA 240
Db 181 TTGGTTCAGACAAATCGGTCGAGGCGGAGCGCATTCGTCTGAAGCGCTGGTCAAGA 240
QY 241 GCGAGCGTATTCCACACCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 300
Db 241 GCGAGCGTATTCCACACCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 300
QY 301 ACACCGAGGCAAGGCAATGGATCCGTCGCGATGAAGATGATGGCGGACGGCTTCTG 360
Db 301 ACACCGAGGCAAGGCAATGGATCCGTCGCGATGAAGATGATGGCGGACGGCTTCTG 360
QY 361 GTGGTGTCTGTCTCCACACCAAGCATGTGTTCTGATGCTCACTGATTGGAATTT 420
Db 361 GTGGTGTCTGTCTCCACACCAAGCATGTGTTCTGATGCTCACTGATTGGAATTT 420
QY 421 GTATCTACCAAGTATCCCTGGAGAGTGAGAGTGTATTCAGTCTATTTATCTTGAT 480
Db 421 GTATCTACCAAGTATCCCTGGAGAGTGAGAGTGTATTCAGTCTATTTATCTTGAT 480
QY 481 GTATCTGCCCTTGTCTCCCTCAGAGTATTCGTTTGTGGATGAGACAAAGTGA 540
Db 481 GTATCTGCCCTTGTCTCCCTCAGAGTATTCGTTTGTGGATGAGACAAAGTGA 540
QY 541 ATAAGAGTGTCTATATACAGCATTCCTGTGTGTGTGTGTGTGTGTGTGTGTGTCA 600
Db 541 ATAAGAGTGTCTATATACAGCATTCCTGTGTGTGTGTGTGTGTGTGTGTGTGTCA 600
QY 601 TGTATCTGTAATTTGATGTCGGATTTCTACTATTTATCAATCTCATTTACTGTGT 660
Db 601 TGTATCTGTAATTTGATGTCGGATTTCTACTATTTATCAATCTCATTTACTGTGT 660

QY 661 GTAAAAA 673
Db 661 GTAAAAA 673
RESULT 2
ABQ28226
ID ABQ28226 standard; DNA; 503 BP.
AC ABQ28226;
DT 12-JUL-2002 (first entry)
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 14917.
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
OS Homo sapiens.
XX WO200218632-A2.
XX PD 07-MAR-2002.
XX PF 01-SEP-2001; 2001WO-EP10074.
XX PR 01-SEP-2000; 2000DE-1043826.
XX PR 05-SEP-2000; 2000DE-1044543.
XX PA (BPIG-) EPIGENOMICS AG.
XX PI Olek A, Piepenbrock C, Berlin X, Guetig D;
XX WPI; 2002-371829/40.
XX PT Determining the degree of cytosine methylation in genomic DNA, useful
XX for diagnosis and prognosis, comprises selective hybridization of
XX amplicons from chemically treated DNA -
XX
XX Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
XX This invention describes a novel method for determining the degree of
XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX genomic sample of DNA. The sample is treated chemically to convert
XX cytosine (C) but not methylated C to uracil, then part of the genomic
XX DNA that contains the target C is amplified to form a labeled amplicon.
XX The amplicon is hybridised to two classes, each with at least one
XX member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
XX and the degree of hybridisation to both classes is determined from the
XX label on the amplicon. From the ratio of labels hybridised to the two
XX classes of oligomers, the degree of methylation is calculated. The method
XX is used: (i) for diagnosis and/or prognosis of side effects of
XX therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
XX of the central nervous, cardiovascular, gastrointestinal and respiratory
XX systems etc.; particularly by detecting mutations or single nucleotide
XX polymorphisms (SNPs); and (ii) for differentiation of cell or tissue
XX types and for investigating cell differentiation. The method allows the
XX methylation status of many C residues to be determined simultaneously.
XX ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
XX method for determining the degree of cytosine methylation described in
XX the disclosure of the invention.
XX
XX Sequence 503 BP; 86 A; 55 C; 171 G; 191 T; 0 other;
XX SQ
Query Match 3.0%; Score 20; DB 24; Length 503;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 504 AGAAGTATTCGTTGTTGT 523
Db 374 AGAAGTATTCGTTGTTGT 393

RESULT 3

ABQ28227/c
 ID ABQ28227 standard; DNA; 503 BP.
 XX
 AC ABQ28227;
 XX
 DT 12-JUL-2002 (first entry)
 XX
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 14818.
 XX
 KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 XX drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200218632-A2.
 XX
 PD 07-MAR-2002.
 XX
 PF 01-SEP-2001; 2001WO-EP10074.
 XX
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PR 05-SEP-2000; 2000DE-1044543.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;
 XX
 DR WPI; 2002-371829/40.
 XX

Determining the degree of cytosine methylation in genomic DNA, useful
 PT for diagnosis and prognosis, comprises selective hybridization of
 PT amplicons from chemically treated DNA

PS Claim 12; 56pp + Sequence Listing; 56pp; German.

CC This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
 CC and the degree of hybridisation to both classes is determined from the
 CC label on the amplicon. From the ratio of labels hybridised to the two
 CC classes of oligomers, the degree of methylation is calculated. The method
 CC is used: (i) for diagnosis and/or prognosis of side effects of
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory
 CC systems etc., particularly by detecting mutations or single nucleotide
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
 CC types and for investigating cell differentiation. The method allows the
 CC methylation status of many C residues to be determined simultaneously.
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
 CC method for determining the degree of cytosine methylation described in
 CC the disclosure of the invention.

XX SQ Sequence 503 BP; 191 A; 171 C; 55 G; 86 T; 0 other;

Query Match 3.0%; Score 20; DB 24; Length 503;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 504 AGAAGTATTCGTTTGTGT 523
 DB 130 AGAAGTATTCGTTTGTGT 111

RESULT 4

ABQ53592
 ID ABQ53592 standard; DNA; 569 BP.

XX
 AC ABQ53592;
 XX
 DT 12-JUL-2002 (first entry)
 XX
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 40183.
 XX
 KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200218632-A2.
 XX
 PD 07-MAR-2002.
 XX
 PF 01-SEP-2001; 2001WO-EP10074.
 XX
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PR 05-SEP-2000; 2000DE-1044543.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;
 XX
 DR WPI; 2002-371829/40.

Determining the degree of cytosine methylation in genomic DNA, useful
 PT for diagnosis and prognosis, comprises selective hybridization of
 PT amplicons from chemically treated DNA

PS Claim 12; 56pp + Sequence Listing; 56pp; German.

CC This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
 CC and the degree of hybridisation to both classes is determined from the
 CC label on the amplicon. From the ratio of labels hybridised to the two
 CC classes of oligomers, the degree of methylation is calculated. The method
 CC is used: (i) for diagnosis and/or prognosis of side effects of
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory
 CC systems etc., particularly by detecting mutations or single nucleotide
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
 CC types and for investigating cell differentiation. The method allows the
 CC methylation status of many C residues to be determined simultaneously.
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
 CC method for determining the degree of cytosine methylation described in
 CC the disclosure of the invention.

XX SQ Sequence 569 BP; 92 A; 61 C; 191 G; 225 T; 0 other;

Query Match 3.0%; Score 20; DB 24; Length 569;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 504 AGAAGTATTCGTTTGTGT 523
 DB 475 AGAAGTATTCGTTTGTGT 494

RESULT 5

ABQ53593/c
 ID ABQ53593 standard; DNA; 569 BP.
 XX

AC ABQ53593;
 XX
 DT 12-JUL-2002 (first entry)
 XX
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 40184.
 XX
 KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 XX drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.
 XX
 OS Homo sapiens.
 XX
 XX WO200218632-A2.
 FN
 XX
 XX 07-MAR-2002.
 PD
 XX
 XX 01-SEP-2001; 2001WO-EP10074.
 PF
 XX
 XX 01-SEP-2000; 2000DE-1043826.
 PR
 XX 05-SEP-2000; 2000DE-1044543.
 PR
 XX (EPIG-) EPIGENOMICS AG.
 XX
 XX Olek A, Piepenbrock C, Berlin K, Guetig D;
 PI
 XX WPI; 2002-371829/40.
 DR
 XX
 XX Determining the degree of cytosine methylation in genomic DNA, useful
 PT for diagnosis and prognosis, comprises selective hybridization of
 PT amplicons from chemically treated DNA
 XX
 XX Claim 12; 56pp + Sequence Listing; 56pp; German.
 PS
 XX
 CC This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
 CC and the degree of hybridisation to both classes is determined from the
 CC label on the amplicon. From the ratio of labels hybridised to the two
 CC classes of oligomers, the degree of methylation is calculated. The method
 CC is used: (i) for diagnosis and/or prognosis of side effects of
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory
 CC systems etc., particularly by detecting mutations of single nucleotide
 CC polymorphisms (SNPs); and (ii) for differentiation of cell or tissue
 CC types and for investigating cell differentiation. The method allows the
 CC methylation status of many C residues to be determined simultaneously.
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
 CC method for determining the degree of cytosine methylation described in
 CC the disclosure of the invention.
 XX
 SQ Sequence 569 BP; 225 A; 191 C; 61 G; 92 T; 0 other;
 Query Match 3.0%; Score 20; DB 24; Length 569;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 504 AGAAGTATTCGTTGTTGT 523
 Db |||||
 95 AGAAGTATTCGTTGTTGT 76
 RESULT 6
 AAS22945
 ID AAS22945 standard; DNA; 1967 BP.
 XX
 AC AAS22945;
 XX
 XX 24-OCT-2001 (first entry)
 DT

XX DNA encoding novel bone marrow polypeptide #39.
 DE
 XX Bone marrow; diagnostic; therapeutic; gene therapy; antigenic;
 KW haematopoiesis; myeloid; lymph cell disorder; tissue regeneration;
 KW wound healing; nutritional supplement; immune disorder;
 KW severe combined immunodeficiency; SCID; ds.
 XX
 OS Homo sapiens.
 XX
 XX WO200157187-A2.
 FN
 XX
 XX 09-AUG-2001.
 PD
 XX
 XX 05-FEB-2001; 2001WO-US03782.
 PF
 XX
 XX 03-FEB-2000; 2000US-0496914.
 PR
 XX 20-JUN-2000; 2000US-0598075.
 PR
 XX 19-JUL-2000; 2000US-0620325.
 PR
 XX 30-NOV-2000; 2000US-0250683;
 XX (HYSB-) HYSEQ INC.
 PA
 XX Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Zhou P, Xue AJ;
 PI Ren F, Drmanac RT;
 PI
 XX WPI; 2001-488875/53.
 DR
 XX P-PSDB; AAU14640.
 DR
 XX
 XX Nucleic acids encoding bone marrow polypeptides, useful in diagnostic
 PT and gene therapy -
 PT
 XX
 XX Claim 1; Page 177; 392pp; English.
 PS
 XX
 CC AAS22907-AAS23099 represent nucleic acids encoding novel bone marrow
 CC polypeptides. The nucleic acids and corresponding proteins may be used
 CC in the prevention, diagnosis and treatment of diseases associated with
 CC inappropriate bone marrow polypeptide expression. For example, to treat
 CC disorders associated with decreased expression by rectifying mutations
 CC or deletions in a patient's genome that affect the activity of the
 CC polypeptides by expressing inactive proteins or to supplement the
 CC patient's own production of the polypeptide. Additionally, the nucleic
 CC acids may be used to produce the polypeptides, by inserting the nucleic
 CC acids into a host cell and culturing the cell to express the protein.
 CC The nucleic acid and its complementary sequences may also be used as DNA
 CC probes in diagnostic assays to detect and quantitate the presence of
 CC similar nucleic acid sequences in samples, and therefore which patients
 CC may be in need of restorative therapy. The proteins may also be used as
 CC antigens in the production of antibodies against bone marrow proteins
 CC and in assays to identify modulators of their expression and activity.
 CC The anti-bone marrow protein antibodies and antagonists may also be used
 CC to down regulate expression and activity. The antibodies may also be used
 CC as diagnostic agents for detecting the presence of the protein in samples
 CC (e.g. by enzyme linked immunosorbent assay (ELISA)). The proteins
 CC may be used to regulate haematopoiesis activity, and consequently in the
 CC treatment of myeloid or lymph cell disorders; in tissue regeneration,
 CC such as wound healing; as a nutritional supplement; and in treatment of
 CC immune disorders such as severe combined immunodeficiency (SCID).
 XX
 SQ Sequence 1967 BP; 456 A; 514 C; 622 G; 375 T; 0 other;
 Query Match 3.0%; Score 20; DB 22; Length 1967;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 618 GATGCTGGATTCCTACTATT 637
 Db |||||
 1616 GATGCTGGATTCCTACTATT 1635
 RESULT 7
 AAF07706/C
 ID AAF07706 standard; cDNA; 525 BP.

XX Frog embryonic gene sequence 09925363.
 DE
 XX
 KW Frog; ss; embryonic development; developmental disorder;
 KW microarray; cell differentiation.
 XX
 OS Xenopus laevis.
 XX
 PN US2002081610-A1.
 XX
 PD 27-JUN-2002.
 XX
 XX 23-JUL-2001; 2001US-0910943.
 PF
 XX 21-JUL-2000; 2000US-219658P.
 PR
 XX (UVRQ) UNIV ROCKEFELLER.
 PA
 XX Hemmati-Brivanlou A, Altmann CR;
 PI
 XX WPI; 2002-628534/67.
 DR
 XX Nucleic acid array containing Xenopus embryonic nucleic acids is useful
 PT to identify genes involved in embryonic development, to identify
 PT different types of embryonic cells, and to diagnose developmental
 PT disorders.
 XX
 PS Claim 1; Page 239; 823pp; English.
 XX
 CC The invention relates to a nucleic acid array, where each coordinate
 CC contains a single nucleic acid species having one of 770 nucleotide
 CC sequences (appearing as ABS76747-ABS77516) a of a Xenopus embryonic gene
 CC product, or its complement or hybridisable fragment of not less than
 CC 20 contiguous nucleotides of one of those sequences. Also included are
 CC detecting differential expression of embryonic genes, comprising:
 CC (a) contacting a nucleic acid array comprising genes expressed in
 CC embryonic but not mature cells with nucleic acids from sample and
 CC control cells; and (b) detecting differential hybridisation of nucleic
 CC acids from the sample cells relative to the control cells; and detecting
 CC defects in development, comprising: (a) contacting nucleic acids from
 CC test cells undergoing development with a nucleic acid array of gene
 CC products known to play a fundamental role in the development process; and
 CC (b) detecting a difference in expression of a fundamental gene in the
 CC sample cells relative to a standard. The invention is useful to identify
 CC genes involved in embryonic development and related processes such as
 CC cell differentiation. This would be useful for diagnosing developmental
 CC disorders and for identifying different types of embryonic cells.
 CC The present sequence is one of the 770 Xenopus embryonic cDNA sequences.
 XX
 SQ Sequence 720 BP; 173 A; 200 C; 166 G; 179 T; 2 other;
 Query Match 2.8%; Score 19; DB 24; Length 720;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 354 GCTTCTGGTGGTGGTGGTGGT 372
 Db 461 GCTTCTGGTGGTGGTGGTGGT 479
 RESULT 10
 AAF72803
 ID AAF72803 standard; DNA; 2057 BP.
 XX
 AC AAF72803;
 XX
 XX 24-APR-2001 (first entry)
 DT
 XX Secreted protein gene #5.
 DE
 XX Secreted protein; human; autoimmune; hyperproliferation;
 KW cardiovascular; cerebrovascular; infection; food; ds.
 XX

OS Homo sapiens.
 XX
 PN WO200107459-A1.
 XX
 PD 01-FEB-2001.
 XX
 XX 20-JUL-2000; 2000WO-US19735.
 PF
 XX 23-JUL-1999; 99US-0145220.
 PR
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX Rosen CA, Ruben SM, Ebner R, Duan RD, Ni J, Soppet DR, Moore PA;
 PI Shi Y, Lafleur DW, Olsen HS, Birse CB, Komatsoulis GA;
 XX WPI; 2001-123261/13.
 DR
 XX New isolated nucleic acid encoding 29 secreted proteins, for
 PT diagnosing, preventing and treating e.g. autoimmune,
 PT hyperproliferative, cardiovascular, and ocular diseases or disorders
 PT and microorganism infections.
 XX
 PS Claim 1; Page 497; 601pp; English.
 XX
 CC The present invention relates to 29 human secreted proteins. The
 CC invention is used to prevent autoimmune diseases e.g. rheumatoid
 CC arthritis, hyperproliferative disorders e.g. neoplasms of the
 CC breast or liver, cardiovascular disorders e.g. cardiac arrest,
 CC cerebrovascular disorders e.g. cerebral ischemia, angiogenesis,
 CC nervous system disorders e.g. Alzheimer's disease, infections
 CC caused by bacteria, viruses and fungi and ocular disorders e.g.
 CC corneal infection. Also used in food preparations.
 XX
 SQ Sequence 2057 BP; 599 A; 500 C; 502 G; 456 T; 0 other;
 Query Match 2.8%; Score 19; DB 22; Length 2057;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 655 CTGTGTGTAATAAAAAA 673
 Db 1903 CTGTGTGTAATAAAAAA 1921
 RESULT 11
 ABL21565/c
 ID ABL21565 standard; DNA; 3974 BP.
 XX
 AC ABL21565;
 XX
 DT 26-MAR-2002 (first entry)
 DE
 XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 16168.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 XX 23-MAR-2001; 2001WO-US09231.
 PP
 XX 23-MAR-2000; 2000US-191637P.
 PR
 XX 11-JUL-2000; 2000US-0614150.
 PR
 XX (PEKE) PE CORP NY.
 PA
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI
 XX WPI; 2001-656860/75.
 DR

XX		New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -
PT		Claim 1; SEQ ID NO 16168; 21pp + Sequence Listing; English.
PS		The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-ABBS72072).
CC		The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pat_sequences..
CC		Sequence 3974 BP; 1151 A; 964 C; 881 G; 978 T; 0 other;
XX		
SQ		
	Query Match	2.8%; Score 19; DB 23; Length 3974;
	Best Local Similarity	100.0%; Pred.No. 33;
	Matches 19; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	356 TTCTGGTGGTGCTGCTGCT	374
DB	348 TTCTGGTGGTGCTGCTGCT	330
RESULT 12		
AAS36235	ID	AAS36235 standard; DNA; 7537 BP.
XX	AC	AAS36235;
XX	DT	(first entry)
XX		Human cardiovascular system antigen genomic DNA SEQ ID No 1735.
XX		Cardiovascular system antigen; human; mouse; rabbit; goat; horse; cat; chicken; sheep; immunosuppressive; antiarthritic; vasotropic; dog; antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective; cerebroprotective; nocotropic; antibacterial; virucide; fungicide; cancer; opthalmological; vulnerrary; gene therapy; autoimmune disease; neoplasm; hyperproliferative disorder; breast; liver; cardiovascular disorder; ds; cerebrovascular disorder; nervous system disorder; bacterial infection; fungal infection; viral infection; ocular disorder; endocrine disorder; gastrointestinal disorder; renal disorder; respiratory disorder; wound healing; skin aging; organ transplantation; tissue regeneration; anti-infertility.
OS	Homo sapiens.	
XX		
PN	WO200155321-A2.	
XX		
PD	02-AUG-2001.	
XX		
PF	17-JAN-2001; 2001MO-US01340.	
XX		
PR	31-JAN-2000; 2000US-0179055.	
PR	04-FEB-2000; 2000US-0180628.	
PR	24-FEB-2000; 2000US-0184664.	
PR	02-MAR-2000; 2000US-0186350.	
PR	16-MAR-2000; 2000US-0189874.	
PR	17-MAR-2000; 2000US-0190076.	
PR	18-APR-2000; 2000US-0198123.	
PR	19-MAY-2000; 2000US-0205515.	
PR	07-JUN-2000; 2000US-0209467.	
PR	28-JUN-2000; 2000US-0214886.	
PR	30-JUN-2000; 2000US-0215135.	
PR	07-JUL-2000; 2000US-0216647.	
PR	07-JUL-2000; 2000US-0216880.	

PR 20-OCT-2000; 2000US-0241926.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249284.
 PR 17-NOV-2000; 2000US-0249285.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 05-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 08-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 12-DEC-2000; 2000US-0254097.
 PR 03-JAN-2001; 2001US-02559678.
 (HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Barash SC, Ruben SM;
 WI; 2001-451930/48.

PT New cardiovascular system related polynucleotides and polypeptides,
 PT useful for diagnosing, treating and/or preventing disorders of the
 PT cardiovascular system -
 PS Claim 1; SEQ ID No 1735; 674bp; English.

CC Sequences AAS35741-AAS36942 represent genomic DNA molecules, which encode
 CC the cardiovascular system antigen polypeptides of the invention.
 CC Cardiovascular system antigens and their associated polynucleotides are
 CC useful in the diagnosis, treatment and prevention of various types of
 CC disorders in e.g. humans, mice, rabbits, goats, horses, cats, dogs,
 CC chickens or sheep. A pathological condition can be determined by
 CC detecting the presence or absence of a mutation in a cardiovascular
 CC system antigen polynucleotide. The treatable disorders include autoimmune
 CC diseases such as rheumatoid arthritis, hyperproliferative disorders such
 CC as neoplasms of the breast or liver, cardiovascular disorders such as

CC cardiac arrest, cerebrovascular disorders such as cerebral ischaemia,
 CC nervous system disorders such as Alzheimer's disease, infections caused
 CC by bacteria, viruses and fungi, ocular disorders such as corneal
 CC infection, endocrine disorders such as premature labour and infertility,
 CC gastrointestinal disorders such as Crohn's disease, renal disorders such
 CC as glomerulonephritis and respiratory disorders such as asthma and
 CC pleurisy. The polypeptides can also be used to aid wound healing, to
 CC prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, to regenerate tissues and in chemotaxis.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX

Query Match 2.8%; Score 19; DB 22; Length 7537;

Best Local Similarity 100.0%; Pred. No. 33;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 655 CTGTGTGTAATAAAAAA 673
 |||||
 DB 5800 CTGTGTGTAATAAAAAA 5818

RESULT 13

AAK65643

ID AAK65643 standard; DNA; 7537 BP.

XX AAK65643;

DT 06-NOV-2001 (first entry)

DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:20455.

KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

XX cytostatic; gene therapy; vaccine; metastasis; ds.

OS Homo sapiens.

XX WO200157182-A2.

XX 09-AUG-2001.

XX 17-JAN-2001; 2001WO-US01354.

XX 31-JAN-2000; 2000US-0179065.

XX 04-FEB-2000; 2000US-0180628.

XX 24-FEB-2000; 2000US-0184664.

XX 02-MAR-2000; 2000US-0186350.

XX 16-MAR-2000; 2000US-0189874.

XX 17-MAR-2000; 2000US-0190076.

XX 18-APR-2000; 2000US-0198123.

XX 19-MAY-2000; 2000US-0205515.

XX 07-JUN-2000; 2000US-0209467.

XX 30-JUN-2000; 2000US-0215135.

XX 07-JUL-2000; 2000US-0216647.

XX 07-JUL-2000; 2000US-0216880.

XX 11-JUL-2000; 2000US-0217487.

XX 11-JUL-2000; 2000US-0217496.

XX 14-JUL-2000; 2000US-0218290.

XX 26-JUL-2000; 2000US-0220963.

XX 26-JUL-2000; 2000US-0220964.

XX 14-AUG-2000; 2000US-0224518.

XX 14-AUG-2000; 2000US-0224519.

XX 14-AUG-2000; 2000US-0225213.

XX 14-AUG-2000; 2000US-0225214.

XX 14-AUG-2000; 2000US-0225266.

XX 14-AUG-2000; 2000US-0225267.

XX 14-AUG-2000; 2000US-0225268.

XX 14-AUG-2000; 2000US-0225270.

XX 14-AUG-2000; 2000US-0225447.

XX 14-AUG-2000; 2000US-0225757.

XX 14-AUG-2000; 2000US-0225758.

PR	14-AUG-2000;	2000US-0225759;
PR	18-AUG-2000;	2000US-022562759;
PR	22-AUG-2000;	2000US-0226681;
PR	22-AUG-2000;	2000US-0226868;
PR	22-AUG-2000;	2000US-0227182;
PR	23-AUG-2000;	2000US-0227009;
PR	30-AUG-2000;	2000US-0228294;
PR	30-AUG-2000;	2000US-0228923;
PR	01-SEP-2000;	2000US-0229287;
PR	01-SEP-2000;	2000US-0229343;
PR	01-SEP-2000;	2000US-0229344;
PR	01-SEP-2000;	2000US-0229345;
PR	05-SEP-2000;	2000US-0229509;
PR	05-SEP-2000;	2000US-0229513;
PR	06-SEP-2000;	2000US-0230437;
PR	06-SEP-2000;	2000US-0230438;
PR	08-SEP-2000;	2000US-0231242;
PR	08-SEP-2000;	2000US-0231243;
PR	08-SEP-2000;	2000US-0231244;
PR	08-SEP-2000;	2000US-0231413;
PR	08-SEP-2000;	2000US-0231414;
PR	08-SEP-2000;	2000US-0232080;
PR	08-SEP-2000;	2000US-0232081;
PR	14-SEP-2000;	2000US-0231968;
PR	14-SEP-2000;	2000US-0232397;
PR	14-SEP-2000;	2000US-0232398;
PR	14-SEP-2000;	2000US-0232399;
PR	14-SEP-2000;	2000US-0232400;
PR	14-SEP-2000;	2000US-0232401;
PR	14-SEP-2000;	2000US-0233063;
PR	14-SEP-2000;	2000US-0233064;
PR	14-SEP-2000;	2000US-0233065;
PR	21-SEP-2000;	2000US-0234223;
PR	21-SEP-2000;	2000US-0234274;
PR	25-SEP-2000;	2000US-0234997;
PR	25-SEP-2000;	2000US-0234998;
PR	25-SEP-2000;	2000US-0235484;
PR	27-SEP-2000;	2000US-0235834;
PR	27-SEP-2000;	2000US-0235836;
PR	29-SEP-2000;	2000US-0236327;
PR	29-SEP-2000;	2000US-0236367;
PR	29-SEP-2000;	2000US-0236368;
PR	29-SEP-2000;	2000US-0236369;
PR	29-SEP-2000;	2000US-0236802;
PR	02-OCT-2000;	2000US-0237037;
PR	02-OCT-2000;	2000US-0237038;
PR	02-OCT-2000;	2000US-0237039;
PR	03-OCT-2000;	2000US-0237040;
PR	13-OCT-2000;	2000US-0239935;
PR	13-OCT-2000;	2000US-0239937;
PR	20-OCT-2000;	2000US-0240826;
PR	20-OCT-2000;	2000US-0241809;
PR	20-OCT-2000;	2000US-0241808;
PR	20-OCT-2000;	2000US-0241826;
PR	01-NOV-2000;	2000US-0244617;
PR	08-NOV-2000;	2000US-0246474;
PR	08-NOV-2000;	2000US-0246475;
PR	08-NOV-2000;	2000US-0246476;
PR	08-NOV-2000;	2000US-0246477;
PR	08-NOV-2000;	2000US-0246478;
PR	08-NOV-2000;	2000US-0246523;
PR	08-NOV-2000;	2000US-0246524;
PR	08-NOV-2000;	2000US-0246525;
PR	08-NOV-2000;	2000US-0246526;
PR	09-NOV-2000;	2000US-0246527;
PR	08-NOV-2000;	2000US-0246532;
PR	08-NOV-2000;	2000US-0246533;
PR	08-NOV-2000;	2000US-0246538;
PR	08-NOV-2000;	2000US-0246610;

PR	08-NOV-2000;	2000US-0245611.1.
PR	08-NOV-2000;	2000US-0245613.
PR	17-NOV-2000;	2000US-0249207.
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249264.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249269.
PR	17-NOV-2000;	2000US-0249299.
PR	17-NOV-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0255678.
XX	(HUMA-)	HUMAN GENOME SCI INQ
PA		
XX		
PA	Rosen CA, Barash SC, Ruben	
PI	WPI; 2001-483426/52.	
XX		
DR		
XX		
DR	Nucleic acids encoding human	
PT	useful for preventing, diagnosing	
PT	metastasis -	
FI		
XX		
PS	Disclosure: SEQ ID NO 20455.	
XX		
XX	AAK54951 to AAK54702 encode	
CC	amino acid sequences given in	
CC	activity, and can be used in	
CC	proteins and polynucleotides	
CC	treatment of diseases associat	
CC	example, they may be used to	
CC	expression by rectifying mutat	
CC	that affect the activity of	
CC	supplement the patients own	
CC	polynucleotides may be used	
CC	the nucleic acids into a host	
CC	protein. (I) proteins and po	
CC	diagnose and treat immune/h	
CC	cancers and cancer metatases	
CC	to AAK87694 represent human	
CC	sequences from the present i	
CC	represent sequences used in	
XX		
SQ	Sequence 7537 BP; 1856 A; 18	
		2.8%;
	Query Watch	Best Local Similarity 100.0%;
	Matches 19; Conservative	
Qy	655 CTGTGCTGTAATAAAAAA	

Db 5800 CTGTGTGTAATAAAAAA 5818

RESULT 14
ABL21564
ID ABL21564 standard; DNA; 15366 BP.
XX
AC ABL21564;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 16165.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PI 11-JUL-2000; 2000US-0614150.
XX
PA (PEXE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
PS Claim 1; SEQ ID NO 16165; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (AB057737-AB072072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 15366 BP; 4370 A; 3215 C; 3333 G; 4448 T; 0 other;
Query Match 2.8%; Score 19; DB 23; Length 15366;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 356 TTCTGGTGGTCTGCTGCT 374
Db 14019 TTCTGGTGGTCTGCTGCT 14037
RESULT 15
ABQ81850/c
ID ABQ81850 standard; DNA; 156638 BP.
XX
AC ABQ81850;
XX
DT 19-NOV-2002 (first entry)
XX
DE Bifidobacterium longum NCC2705 related nucleotide sequence SEQ ID:1106.
XX
KW Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;

KW antidiarrhetic; antibacterial; inhibitor of Salmonella; detection;
KW identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;
KW rotavirus; food composition; pharmaceutical composition; gene; ds.
XX
OS Bifidobacterium longum.
OS Synthetic.
XX
PN EP1227152-A1.
XX
PD 31-JUL-2002.
XX
PF 30-JAN-2001; 2001EP-0102050.
XX
PR 30-JAN-2001; 2001EP-0102050.
XX
PA (NEST) SOC PROD NESTLE SA.
XX
WPI; 2002-668397/72.
XX
PT Novel polynucleotide comprising Bifidobacterium genome sequence useful
PT as a probe or primer for detecting and/or identifying Bifidobacterium
PT longum in a biological sample.
XX
PS Disclosure; SEQ ID 1106; 80pp; English.
XX
CC The present invention describes a polynucleotide (I) comprising a
CC sequence of a Bifidobacterium genome selected from the nucleotide
CC sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at
CC least 90% identity or which hybridises with the sequences given in
CC ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding
CC a fusion protein, comprising a sequence selected from 1097 sequences
CC given in ABP65258 to ABP66354 ligated in frame to a polynucleotide
CC encoding a heterologous polypeptide. (I) has an antidiarrhetic and
CC antibacterial activities, and can be used as an inhibitor of Salmonella.
CC (I) (which is a probe) is useful for the detection and/or identification
CC of Bifidobacterium longum in a biological sample. A carrier containing
CC the lactic acid bacterium Bifidobacterium longum NCC2705 (NCM 1-2618/
CC can be used for preventing and/or treating diarrhoea brought about by
CC pathogenic bacteria and/or rotavirus. The carrier is a food composition
CC selected from milk, yogurt, curd, cheese, fermented milks, milk based
CC fermented products, ice-creams, fermented cereal based products, milk
CC based powders, infant formula, pet food or a pharmaceutical composition
CC selected from tablets, liquid bacterial suspensions, dried oral
CC supplement, wet oral supplement, dry tube feeding or wet tube feeding.
CC (I) is useful in DNA arrays or chips to carry out analysis of the
CC expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent
CC Bifidobacterium related nucleotide sequences given in the Sequence
CC Listing from the present invention but not mentioned further within the
CC specification.
CC N.B. The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied by the
CC European Patent Office.
XX
SQ Sequence 156638 BP; 32098 A; 46491 C; 46415 G; 31634 T; 0 other;
Query Match 2.8%; Score 19; DB 24; Length 156638;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 331 CGATGAGATGATGCGGG 349
Db 2826 CGATGAGATGATGCGGG 2808
Search completed: November 14, 2003, 10:28:49
Job time : 227 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: November 14, 2003, 09:21:42 ; Search time 1934 Seconds
(without alignments)
14235.873 Million cell updates/sec

Title: US-09-868-025-1
Perfect score: 673
Sequence: 1 ttacactgctgtcggat.....actgtgtataaaaaaaa 673

Scoring table: OLIGO NUC
Gapop 60.0, Gapext 60.0

Searched: 2888711 seqs, 20454813386 residues

Word size: 0
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_in.*
5: gb_ov.*
6: gb_ov.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vt.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_mu.*
21: em_mu.*
22: em_ov.*
23: em_ov.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vt.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rdd.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgc_hum.*
40: em_htgc_mus.*
41: em_htgc_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	673	100.0	673	8	AF192975	AF192975 Oryza sat
2	528	78.5	123673	8	AP002837	AP002837 Oryza sat
3	45	6.7	1164	6	AX653397	AX653397 Sequence
4	45	6.7	80327	2	AP003955	AP003955 Oryza sat
5	45	6.7	191580	8	AP003847	AP003847 Oryza sat
6	22	3.3	108859	4	AL773560	AL773560 Pig DNA s
7	21	3.1	121496	9	AL391123	AL391123 Human DNA
8	21	3.1	145257	2	AC108178	AC108178 Felle cat
9	21	3.1	168227	9	AC021868	AC021868 Homo sapi
10	21	3.1	196817	9	AC019209	AC019209 Homo sapi
11	21	3.1	205592	2	AC044811	AC044811 Homo sapi
12	21	3.1	211289	2	AC109203	AC109203 Mus muscu
13	21	3.1	217807	2	AC117034	AC117034 Rattus no
14	21	3.1	248358	2	AC122944	AC122944 Rattus no
15	21	3.1	254489	2	AC098275	AC098275 Rattus no
16	20	3.0	379	9	HUMMANR29	M93220 Human macro
17	20	3.0	927	8	AY144949	AY144949 Saccharom
18	20	3.0	2624	9	AK090818	AK090818 Homo sapi
19	20	3.0	11308	2	AC135709	AC135709 Rattus no
20	20	3.0	33404	9	BX255924	BX255924 Human DNA
21	20	3.0	70199	8	NCB10N12	NCB10N12 Neurospor
22	20	3.0	77702	9	AC034241	AC034241 Homo sapi
23	20	3.0	102507	9	HSJ324N14	AL109938 Human DNA
24	20	3.0	129694	2	AC108181	AC108181 Felle cat
25	20	3.0	131345	9	AC069023	AC069023 Homo sapi
26	20	3.0	135545	2	AC025184	AC025184 Homo sapi
27	20	3.0	135800	2	AP523316	AP523316 Canis fam
28	20	3.0	137473	2	AC015944	AC015944 Homo sapi
29	20	3.0	144034	2	AC069244	AC069244 Homo sapi
30	20	3.0	145122	9	AC012640	AC012640 Homo sapi
31	20	3.0	150618	2	AP005424	AP005424 Oryza sat
32	20	3.0	158335	2	AC025393	AC025393 Homo sapi
33	20	3.0	160822	2	AC020920	AC020920 Homo sapi
34	20	3.0	162308	2	AP005637	AP005637 Oryza sat
35	20	3.0	166895	2	AC023415	AC023415 Homo sapi
36	20	3.0	168513	2	AC011982	AC011982 Homo sapi
37	20	3.0	168837	9	AL383504	AL383504 Human DNA
38	20	3.0	169089	9	AC055725	AC055725 Homo sapi
39	20	3.0	171491	2	AC022606	AC022606 Homo sapi
40	20	3.0	173117	2	AL157885	AL157885 Homo sapi
41	20	3.0	174054	2	AC131148	AC131148 Homo sapi
42	20	3.0	179558	2	AP005838	AP005838 Oryza sat
43	20	3.0	180397	2	AP003615	AP003615 Oryza sat
44	20	3.0	181892	2	BX510931	BX510931 Danio rer
45	20	3.0	197352	2	AC116264	AC116264 Rattus no

ALIGNMENTS

RESULT 1
AF192975
LOCUS AF192975
DEFINITION Oryza sativa unknown gene.
ACCESSION AF192975
VERSION AF192975.1 GI:6289051
KEYWORDS Oryza sativa (indica cultivar-group)
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzoae; Oryza.
1 (bases 1 to 673)
Patell,V.M., Mathai,C.A., Divya,C. and Ashok,M.
Oryza sativa Variety IR64 (CDNA clone AGT9AL-11 from 7 days old

seedling)
Unpublished
2 (bases 1 to 673)
Patel1,V.M., Mathai,C.A., Divya,C. and Ashok,M.
Direct Submission
Submitted (08-OCT-1999) Plant Genome Biology Department, Avesthagen
Graine Technologies, P.Box 5091, Cubbon Park G.P.O., Bangalore,
Karnataka 560001, India
Location/Qualifiers
1. 673
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="IR64"
/db_xref="taxon:39946"
/clone="AGTSAL-11"
/issue type="7 day old seedling"
/complement(1..431)
/product="unknown"
/function="role in salinity stress response"
/codon_start=1
/product="unknown"
/protein_id="AA06789.1"
/db_xref="GI:6289052"
/translation="MVDNFPISQSETHAWCNSSTTRPSRHHHRERIPCLAGV
TAICSLVHVSHGSGISVALCSQCLNLRPPDCLTNNGGCGECHSLGHVDRF
PQNSNWSGNDSCPLRGEFLVLTTPHKEFAIRAGKV"
BASE COUNT 163 a 145 c 170 g 195 t
ORIGIN
Query Match 100.0%; Score 673; DB 8; Length 673;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 673; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTATCTGCTGCTCGGATGACCAAACTCCATCTTGGAGTGTGGCGTGAGCACCA 60
DB 1 TTATCTGCTGCTCGGATGACCAAACTCCATCTTGGAGTGTGGCGTGAGCACCA 60
QY 61 AGAATTCCTCCCTCAGTGTTTGAGCTGTCCATCGCGCTGTTCCTCATTCGATGC 120
DB 61 AGAATTCCTCCCTCAGTGTTTGAGCTGTCCATCGCGCTGTTCCTCATTCGATGC 120
QY 121 TGAGGAAATCTGTCAACATGCCAAGATGCTCCATGCGCATTCACCATAGCAGCTCCATTG 180
DB 121 TGAGGAAATCTGTCAACATGCCAAGATGCTCCATGCGCATTCACCATAGCAGCTCCATTG 180
QY 181 TTGGTCAGACATCGGTCGAGCGGCGGCGCATTCGTCGAGGCACTGGCTCCAAAGA 240
DB 181 TTGGTCAGACATCGGTCGAGCGGCGGCGCATTCGTCGAGGCACTGGCTCCAAAGA 240
QY 241 GCGACGCTGATTCACACCGCTGGTGACATGTATCCAAAGACTCCAAATTCGAGTG 300
DB 241 GCGACGCTGATTCACACCGCTGGTGACATGTATCCAAAGACTCCAAATTCGAGTG 300
QY 301 ACACCGGCGGAGGATCGGATCGGTCGATGAGATGATGCGGCGGCTTCG 360
DB 301 ACACCGGCGGAGGATCGGATCGGTCGATGAGATGATGCGGCGGCTTCG 360
QY 361 GTGGTCTGCTCTCCAAACCAAGCATGTGTTTCGATGTCATCTGATTTGGAATTT 420
DB 361 GTGGTCTGCTCTCCAAACCAAGCATGTGTTTCGATGTCATCTGATTTGGAATTT 420
QY 421 GTATCTACAGTATCCCTGAGAGTGAGAGTTGATTTGATGCTCTATTTCTTGTGAT 480
DB 421 GTATCTACAGTATCCCTGAGAGTGAGAGTTGATTTGATGCTCTATTTCTTGTGAT 480
QY 481 GTAATTCCTTCTGTCCTCAGAGTATTCGTTGTTTGGGATGAGCAAGTGG 540
DB 481 GTAATTCCTTCTGTCCTCAGAGTATTCGTTGTTTGGGATGAGCAAGTGG 540
QY 541 ATAGAGTGCTACTATATACCATCATCTCTGTTTAAAGTTTGCAGTTCTGCAATTCA 600
DB 541 ATAGAGTGCTACTATATACCATCATCTCTGTTTAAAGTTTGCAGTTCTGCAATTCA 600

QY 601 TGTATCTGTAATTTGATGCTGGATTTCTACTATTTATCAATCGCTATTACTGTGT 660
DB 601 TGTATCTGTAATTTGATGCTGGATTTCTACTATTTATCAATCGCTATTACTGTGT 660
QY 661 GTAAAAAATAAAAA 673
DB 661 GTAAAAAATAAAAA 673
RESULT 2
AP002837/2 123673 bp DNA linear PLN 26-OCT-2002
LOCUS Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 6,
DEFINITION BAC clone:OSJNBa0019F11.
ACCESSION AP002837
VERSION AP002837.2 GI:24413940
KEYWORDS Oryza sativa (japonica cultivar-group)
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE 1
Sasaki,T., Matsumoto,T. and Yamamoto,K.
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, BAC
clone:OSJNBa0019F11
Published Only in Database (2000)
2 (bases 1 to 123673)
Sasaki,T., Matsumoto,T. and Yamamoto,K.
Direct Submission
Submitted (03-AUG-2000) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program, Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasakikenias.aifrc.go.jp, URL:http://rpg.dna.aifrc.go.jp/
Tel:81-298-387441, Fax:81-298-38-7468)
On Oct 26, 2002 this sequence version replaced gi:9711842.
Genes were predicted from the integrated results of the following:
GENSCAN (http://ccr-081.mit.edu/GENSCAN.html), PCENESH
(http://www.softberry.com/), GeneMark hm
(http://opal.biology.gatech.edu/GeneMark/), Glimmer
(http://tigr.org/tigr/tdb/glimmer/gimr_form.html), RiceHM
(http://rpg.dna.aifrc.go.jp/RiceHM/), SplicePredictor
(http://bioinformatics.iastate.edu/cgi-bin/sp.cgi), BLASTN and
BLASTX. The genomic sequence was searched against NCBI Nonredundant
protein database, nr
(ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at
RGP. Protein homologies of the coding regions were searched against
NCBI Nonredundant Protein database with BLASTP. ESTs represent the
identified cDNA sequences using BLASTN with the corresponding DBJ
accession no. and RGP clone ID.
A gene with identity or significant homology to a protein is
classified based on the protein name to indicate the homology level
such as same name, 'putative-' and '-like protein'. A gene without
significant homology to any protein but with EST homology (covering
almost the entire length of partial sequence) is classified as an
'unknown' protein. A gene predicted by two or more gene prediction
programs is classified as a 'hypothetical' protein according to
IRGSP standard. A gene predicted by a single gene prediction
program is also classified as a probable 'hypothetical' protein and
is included as a miscellaneous feature of the sequence.
The orientation of the sequence is from -21M13 to M13rev of the
BAC clone. This sequence of OSJNBa0019F11 clone has an overlap with
POS41H01 (DBJ:AP001389) clone at the 5' end. Detailed information
on overlap and assembly quality together with annotation of this
entry is available at
http://rpg.dna.aifrc.go.jp/GenomeSeq.html.
FEATURES
source
1. 123673
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"

```

gene
CDS
/gene="OSUNBa0019F11.5". .21807, 22183. .22256, 23780. (.23875%)
/complement(join(21489..21807,22183..22256,23780..23875))
/gene="OSUNBa0019F11.5". .21807, 22183. .22256, 23780. (.23875%)
/note="contains ESTs AU082454(S3638), D41267(S3638)
unknown protein"
/codon_start=1
/protein_id="BAC22197.1"
/db_xref="GI:24413945"
/translation="MSRKVVVCAVVGFLGVLTAALGFAGAEAGTRYKFLWKQGLVAVK
VGVDSVILALQRATITIAFLLLTGAALNDORGEENYFGSPCYVVKVPVSGGVAL
SLASVAUAIYYVALSSAKSPFNWQQNGIMGPVIPPQSSEPVPHEDTYNRQL
FP"
complement(join(24784..24822,25076..25213))
/gene="OSUNBa0019F11.6". .24822, 25076. .25213))
/complement(join(24784..24822,25076..25213))
/gene="OSUNBa0019F11.6". .24822, 25076. .25213))
/note="hypothetical protein
predicted by GENSCAN etc."
/codon_start=1
/protein_id="BAC22198.1"
/db_xref="GI:24413946"
/translations="MRAAKTPPAAGDGHGGDOLGAGDVHSGAVRHRKHOOORRRG
AHLKKSLKDNGSLS"
complement(26534..26842)
/gene="OSUNBa0019F11.7". .26842)
/complement(26534..26842)
/gene="OSUNBa0019F11.7". .26842)
/note="hypothetical protein
predicted by FGENSEH etc."
/codon_start=1
/protein_id="BAC22199.1"
/db_xref="GI:24413947"
/translation="WMSSSRPPSAVAVVVAVAPARAATAAAQAQATAASARAGA
VATAGSWGRCCHAGEKKMEKEKEEESGGDGGRMTCGFLCDFGDVNRRTVGAVI
"
30877..31350
/gene="OSUNBa0019F11.8". .31350
30877..31350
/gene="OSUNBa0019F11.8". .31350
/note="hypothetical protein
predicted by Glimmer-M etc."
/codon_start=1
/protein_id="BAC22200.1"
/db_xref="GI:24413948"
/translation="MTWSVSGMAVHAHVGMWAAPAALFVLALTATTADASSSQGPCRAY
GGGGGATPFMRAGAAALVTATAQAATAAANAANGARPUSGFALVAFAPAGAYTAN
VLSLIVPVVAAGCVGAGYKMLVAHYVVDLIPLIFVLGAGFVALINQHLE"
join(31340..34033,34110..34249)
/gene="OSUNBa0019F11.9". .34249)
join(31340..34033,34110..34249)
/gene="OSUNBa0019F11.9". .34249)
/note="contains ESTs C73924(E21102), C99694(E21102)
unknown protein"
/codon_start=1
/protein_id="BAC22201.1"
/db_xref="GI:24413949"
/translation="WTAGVIAGSVLGVSFAIAYLCDTFVSDKKAFGGSIPTKVSDKEWMM
QATNKFQAWPRTAGPPVIMNPISRNFIKST"
complement(35938..37863)
/gene="OSUNBa0019F11.10". .37863)
complement(35938..37863)
/gene="OSUNBa0019F11.10". .37863)

gene
CDS
78.5%; Score 528; DB 8; Length 123673;
set Local Similarity 99.7%; Pred. No. 6.8e-289;
atches 628; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

35 TCTTGCGGTGCCTGTCAGCACACCAGAAATTCCTCCCTCAGTGGTTTGCAGCTGTCCA 94
|||||

```

Query Match 78.5%; Score 528; DB 8; Length 123673;
Best Local Similarity 99.7%; Pred. No. 6.8e-288;
Matches 628; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

35 TCCTGGGGTGGCGTGCAGACACCAAGAAATTCCTCCCTCAGTGGTTTGACGTGTCCA 94

57471	Db	TC	TG	GGG	TG	GG	CG	AG	CA	CA	CA	AA	AA	AA	TT	CC	CC	CT	CAG	TGG	TTT	G	CA	G	TG	CC	A	57411										
95	Qy	TG	CG	CG	TG	TT	CC	AT	T	CA	TT	TG	GA	AT	T	CT	GA	AT	TC	GA	AA	AT	CT	CA	AT	CC	CA	AG	AT	CG	CA	T	154					
57411	Db	TG	CG	CG	TG	TT	CC	AT	T	CA	TT	TG	GA	AT	T	CT	GA	AT	TC	GA	AA	AT	CT	CA	AT	CC	CA	AG	AT	CG	CA	T	57352					
155	Qy	GG	AT	T	C	A	C	A	T	A	G	A	C	C	T	C	A	T	T	T	G	G	T	C	A	G	A	A	T	C	G	G	T	214				
57351	Db	GG	AT	T	C	A	C	A	T	A	G	A	C	C	T	C	A	T	T	T	G	G	T	C	A	G	A	A	T	C	G	G	T	57292				
215	Qy	TG	CT	CT	GA	AG	G	C	A	C	T	GG	T	G	C	A	A	A	G	A	G	C	A	C	T	GA	T	CC	A	C	C	G	T	274				
57291	Db	TG	CT	CT	GA	AG	G	C	A	C	T	GG	T	G	C	A	A	A	G	A	G	C	A	C	T	GA	T	CC	A	C	C	G	T	57232				
275	Qy	TG	CA	AA	CA	AG	A	C	T	G	C	AA	AT	TG	CA	GT	GA	CA	CC	GA	GG	CA	AG	GA	T	GG	GA	T	CC	GC	T	CG	CA	T	334			
57231	Db	TG	CA	AA	CA	AG	A	C	T	G	C	AA	AT	TG	CA	GT	GA	CA	CC	GA	GG	CA	AG	GA	T	GG	GA	T	CC	GC	T	CG	CA	T	57172			
335	Qy	GA	A	G	A	T	G	T	G	G	G	A	C	G	G	C	T	T	C	G	G	T	G	T	G	T	C	T	CC	A	C	A	C	A	T	394		
57171	Db	GA	A	G	A	T	G	T	G	G	A	C	G	G	C	T	T	C	G	G	T	G	T	G	T	C	T	CC	A	C	A	C	A	T	57112			
395	Qy	CT	GA	T	GC	T	C	A	C	T	GA	T	TG	AA	AT	TG	AT	CT	T	A	C	A	G	AT	CC	T	CG	A	G	TG	G	A	G	AT	454			
57111	Db	CT	GA	T	GC	T	C	A	C	T	GA	T	TG	AA	AT	TG	AT	CT	T	A	C	A	G	AT	CC	T	CG	A	G	TG	G	A	G	AT	57052			
455	Qy	AT	A	T	T	C	A	G	T	C	A	T	A	T	T	A	T	CT	T	G	T	G	A	T	AA	T	T	GC	TT	T	G	C	CC	T	CA	G	AT	514
57051	Db	AT	A	T	T	C	A	G	T	C	A	T	A	T	T	A	T	CT	T	G	T	G	A	T	AA	T	T	GC	TT	T	G	C	CC	T	CA	G	AT	56992
515	Qy	TT	T	G	TT	T	T	G	G	A	T	G	A	C	A	G	T	GG	A	T	A	A	G	A	G	T	CT	CT	A	T	A	T	A	C	A	G	AT	574
56991	Db	TT	T	G	TT	T	T	G	G	A	T	G	A	C	A	G	T	GG	A	T	A	A	G	A	G	T	CT	CT	A	T	A	T	A	C	A	G	AT	56932
575	Qy	GT	T	A	G	TT	T	T	G	CC	A	G	T	CT	CC	A	G	T	C	A	T	GT	AT	CT	GT	AA	AT	T	T	GA	T	AT	GC	T	GG	AT	CT	634
56931	Db	GT	T	A	G	TT	T	T	G	CC	A	G	T	CT	CC	A	G	T	C	A	T	GT	AT	CT	GT	AA	AT	T	T	GA	T	AT	GC	T	GG	AT	CT	56872
635	Qy	AT	T	A	T	C	A	A	T	CG	T	C	A	T	T	A	C	T	CT																			

RESULT 3
 AX653397
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Oryza sativa
 Oryza sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ebrhartoideae; Oryzeae; Oryza.
 REFERENCE
 1 Chang,H.S., Chen,W., Cooper,B., Glazebrook,J., Goff,S.A., Hou,Y.M.,
 Katagiri,F., Quan,S., Tao,Y., Whitam,S., Xie,Z., Zhu,T. and Zou,G.
 TITLE
 Plant genes involved in defense against pathogens
 JOURNAL
 Patent: WO 0300898-A 3267 03-JAN-2003;
 Syngenta Participations AG (CH)
 FEATURES
 Location/Qualifiers
 1..1164
 /organism="Oryza sativa"
 /mol_type="genomic DNA"
 /db_xref="taxon:4530"
 294 a -286 c 297 g 287 t
 BASE COUNT 294 a -286 c 297 g 287 t
 ORIGIN
 Query Match 6.7%; Score 45; DB 6; Length 1164;
 Best Local Similarity 100.0%; Pred. No. 1.5e-13;
 Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	Db	135	AAACATGCCCAAGACTGCGCATTCACCATAGCAGCTCCATT	179
RESULT 4	AP003955	AP003955	80327 bp	linear
LOCUS	DEFINITION	Oryza sativa (japonica cultivar-group) chromosome 7		
		QJ1425.E04, *** SEQUENCING IN PROGRESS ***		
ACCESSION	VERSION	AP003955		
KEYWORDS	HTG; HTGS PHASE2.	AP003955.1	GI:15021925	
SOURCE	ORGANISM	Oryza sativa (japonica cultivar-group)		
		Oryza sativa (japonica cultivar-group)		
		Eukaryota; Viridiplantae; Streptophyta; Embryophyta		
		Spermatophyta; Magnoliophyta; Liliopsida; Poales; Oryza		
		Ehrhartoideae; Oryzaceae; Oryza		
REFERENCE		1	Sasaki, T., Matsumoto, T. and Yamamoto, K.	
AUTHORS	TITLE	Oryza sativa nipponbare (GAS) genomic DNA, chromosome 7		
		clone:QJ1425.E04		
JOURNAL	REFERENCE	2	Published Only in Database (2001)	
AUTHORS		2 (bases 1 to 80327)		
TITLE		Sasaki, T., Matsumoto, T. and Yamamoto, K.		
JOURNAL		Submitted (25-JUL-2001) Takuji Sasaki, National Institute of Advanced Industrial Science and Technology, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8602, Japan		
		(B-mail:tsasa@nias.affrc.go.jp, URL:http://rgp.dna.nig.ac.jp)		
		Tel:81-298-38-7441, Fax:81-298-38-7468		
COMMENT		The nucleotide sequence of this BAC clone was generated by combining Monsanto and RGP-Japan sequencing data.		
		NOTE: It currently consists of 1 contigs. Gaps between contigs are represented as runs of N. The order of the pieces of the sequence will be replaced by the finished sequence as soon as it is available.		
		* NOTE: This is a 'working draft' sequence.		
		* This sequence will be replaced		
		* by the finished sequence as soon as it is available.		
		* the accession number will be preserved.		
FEATURES	source	Location/Qualifiers		
		1..80327		
		/organism="Oryza sativa (japonica cultivar-group)"		
		/mol_type="genomic DNA"		
		/cultivar="Nipponbare"		
		/db_xref="taxon:39947"		
		/chromosome="7"		
		/clone="QJ1425.E04"		
BASE COUNT	23838	a 16633 c 16495 g 23361 t		
ORIGIN				
Query Match	6.7%	Score 45;	DB 2;	Length 80327
Best Local Similarity	100.0%	Pred. No. 1.1e-13;		
Matches	45;	Conservative 0;	Mismatches 0;	Indels
QY	135	AAACATGCCCAAGACTGCGCATTCACCATAGCAGCTCCATT	179	
Db	33433	AAACATGCCCAAGACTGCGCATTCACCATAGCAGCTCCATT	3344	
RESULT 5	AP003847/c	AP003847	191580 bp	linear
LOCUS	DEFINITION	Oryza sativa (japonica cultivar-group) genomic DNA, BAC clone:QJ1714_H10.		
ACCESSION	VERSION	AP003847		
KEYWORDS		AP003847.3	GI:22831069	
SOURCE	ORGANISM	Oryza sativa (japonica cultivar-group)		
		Oryza sativa (japonica cultivar-group)		

Bakaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE

AUTHORS

1 Sasaki, T., Matsumoto, T. and Yamamoto, K.

Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC

clone: OJ1714_H10

Published Only in Database (2001)

2 (bases 1 to 191580)

Submitted (03-JUL-2001) Takuji Sasaki, National Institute of

Direct Submission

Sasaki, T., Matsumoto, T. and Yamamoto, K.

Agrobiological Sciences, Rice Genome Research Program; Kannondai

2-1-2, Tsukuba, Ibaraki 305-8602, Japan

(E-mail: tsasaki@nias.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/,

Tel: 81-298-3877441, Fax: 81-298-38-7468)

On Sep 13, 2002 this sequence version replaced gi:21912962.

Genes were predicted from the integrated results of the following:

GENSCAN (<http://CCR-081.mit.edu/GENSCAN.html>), FGENESH

(<http://www.softberry.com/>), GeneMark.hmm

(<http://opal.biology.gatech.edu/GeneMark/>), GlimmerM

(http://www.tigr.org/tdb/glimmerm/glm_form.html), RiceHMM

(<http://rgp.dna.affrc.go.jp/RiceHMM/>), SplicePredictor

(<http://bioinformatics.lastate.edu/cgi-bin/sp.cgi>), BLASTN and

BLASTX. The genomic sequence was searched against NCBI NonRedundant

Protein database, nr

(<http://ncbi.nlm.nih.gov/blast/db>) and the cDNA sequence database at

RGF. Protein homologies of the coding regions were searched against

NCBI NonRedundant Protein database with BLASTP. ESTs represent the

identified cDNA sequences using BLASTN with the corresponding DBJ

accession no. and RGP clone ID.

A gene with identity or significant homology to a protein is

classified based on the protein name to indicate the homology level

such as same name, 'putative-' and '-like protein'. A gene without

significant homology to any protein but with EST homology (covering

almost the entire length of partial sequence) is classified as an

'unknown' protein. A gene predicted by two or more gene prediction

programs is classified as a 'hypothetical' protein according to

IRGSP standard. A gene predicted by a single gene prediction

program is also classified as a probable 'hypothetical' protein and

is included as a miscellaneous feature of the sequence.

The orientation of the sequence is from M3rev to -21M13 of the BAC

clone. This sequence of OJ1714_H10 clone has an overlap with

P0455F03 (DBJ:AP005454) clone at the 5' end and with P0428B12

(DBJ:AP004664) clone at the 3' end. The nucleotide sequence of

this BAC clone was generated by combining Monsanto and RGP-Japan

sequencing data. Detailed information on overlap and assembly

quality together with annotation of this entry is available at

<http://rgp.dna.affrc.go.jp/GenomeSeq.html>.

FEATURES

Source

1. 191580

/organism="Oryza sativa (japonica cultivar-group)"

/mol_type="Genomic DNA"

/cultivar="Nipponbare"

/db_xref="taxon:39947"

/chromosomes="7"

/clone="OJ1714_H10"

2. 1915

/gene="OJ1714_H10.1"

2. 1915

/gene="OJ1714_H10.1"

/note="probably inactive due to stop codon(s) in CDS

pseudogene, putative transposase"

/pseudo

/complement(join(2550. .2727,3541. .3596))

/gene="OJ1714_H10.2"

/complement(join(2550. .2727,3541. .3596))

/gene="OJ1714_H10.2"

/note="This category is not included in IRGSP standard.

hypothetical ORF

predicted by GENSCAN"

/complement(join(8831. .8932,9008. .9055,9332. .9475))

/gene="OJ1714_H10.3"

misc_feature

gene

misc_feature

gene

CDS

gene

misc_feature

gene

CDS

gene

CDS

gene

misc_feature

gene

CDS

/complement(join(8831. .8932,9008. .9055,9332. .9475))
/gene="OJ1714_H10.3"
/note="This category is not included in IRGSP standard.
hypothetical ORF
predicted by GeneMark.hmm"
/complement(9646. .10071)
/gene="OJ1714_H10.4"
/complement(9646. .10071)
/gene="OJ1714_H10.4"
/note="This category is not included in IRGSP standard.
hypothetical ORF
predicted by GlimmerM"
/join(11092. .11178,12933. .13856)
/gene="OJ1714_H10.5"
/join(11092. .11178,12933. .13856)
/gene="OJ1714_H10.5"
/note="contains ESTs C19982(E11335), AU075687(E11335)
unknown protein"
/codon_start=1
/protein_id="BAC15932.1"
/db_xref="GI:22831070"
/translation="MAKKKPIVPMDSIGDEPEAIPNPMTRRERERSSGGG
RVRSSRLVYVTPADMDVTHTLFEVSCSYRDMETGRTTSVAGEEAVRAHQVTOG
VRRRLRGSGPGSHRRRAGARVPPRRRLRGVLAAGGVLPERRVPRPQGA
PPAVQGGDTNDLKLRLSLPSFAMLVFLPDDRGRSIGPARQDHLLAGVRRPPAGV
RPRQESDDDLGSLYDGYPPPLKLVDFVADHLFAFFIVEERLQSVFAGHVLDPSE
EV"
/join(14565. .14684,15475. .15609,16934. .16990)
/gene="OJ1714_H10.6"
/join(14565. .14684,15475. .15609,16934. .16990)
/gene="OJ1714_H10.6"
/note="This category is not included in IRGSP standard.
hypothetical ORF
predicted by GENSCAN"
19647. .20096
/gene="OJ1714_H10.7"
19647. .20096
/gene="OJ1714_H10.7"
/note="hypothetical protein
predicted by GlimmerM etc."
/codon_start=1
/protein_id="BAC15933.1"
/db_xref="GI:22831071"
/translation="MQLTRRPFCKIQNLNRRGRRLRLWHEETRPATGRLGVV
RGQWRLARSDAPARLARLRDEAPVTEAGTMRGGAAGGAGGADYGARRLAG
GCHRCSPICQQLDSDGRSGRQHGPGAGGRVVKIQPLSRAGNDDA"
/join(22191. .22749,22910. .22914)
/gene="OJ1714_H10.8"
/join(22191. .22749,22910. .22914)
/gene="OJ1714_H10.8"
/note="hypothetical protein
predicted by GlimmerM etc."
/codon_start=1
/protein_id="BAC15934.1"
/db_xref="GI:22831072"
/translation="MYKPOLCLRRSNVTSRSIPTPSRLMAGAGNDPFFVDGGG
SSKTSNGARAVICLVCAVFLMAGITQCKRVFPAPFPAPPPQRDDDDG
IVVAGIDEATLQALFLVYGEARTAGTCACVCLSYGGGVLRALPQCGHLFHRDCI
FTLRRPTCPVCRAPPAPPADVLS"
/complement(join(24053. .24245,24520. .24552,24656. .24915))
/gene="OJ1714_H10.9"
/complement(join(24053. .24245,24520. .24552,24656. .24915))
/gene="OJ1714_H10.9"
/note="This category is not included in IRGSP standard.
hypothetical ORF
predicted by GENSCAN"
/complement(join(25444. .26085,26479. .26805,26890. .27060))
/gene="OJ1714_H10.10"
/complement(join(25444. .26085,26479. .26805,26890. .27060))
/gene="OJ1714_H10.10"
/note="contains ESTs AU031690(R0565), D33912(R0565)
similar to Oryza sativa chromosome10, OSJNBa0056G1.7

On Oct 3, 2002 this sequence version replaced gi:23395449.

----- Genome Center
Center: Wellcome Trust Sanger Institute
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phased quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; SW: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep XX-514B12 is from a Pig Library
VECTOR: pBeloBAC11

FEATURES
source
1..108859
/location=Qualifiers
/organism="Sus scrofa"
/mol_type="genomic DNA"
/db_xref="taxon:9823"
/clone="XX-514B12"
/clone_lib="SBAB"

BASE COUNT 23091 a 30901 c 31149 g 23718 t
ORIGIN

Query Match 3.3%; Score 22; DB 4; Length 108859;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 TGGTGTCTGCTCCACACCA 383
Db 86922 TGGTGTCTGCTCCACACCA 86901

RESULT 7
AL391123/c
LOCUS 121496 bp DNA linear PRI 25-JUL-2001
DEFINITION Human DNA sequence from clone Rpl1-184118 on chromosome 13, complete sequence.
ACCESSION AL391123
VERSION AL391123.19 GI:15022222
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
BLAKE, S.
Direct Submission
Submitted (25-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT
On Jul 26, 2001 this sequence version replaced gi:13872368.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate

unknown protein"
/codon_start=1
/protein_id="BAC15935.1"
/db_xref="GI:22831073"
/translation="MFTWHQSPNMEFSSFLAESSENLEFHPNVERCPFLRNNG
ATTYFSFALPVAANGNGPFDGPGFDSAPKLFHGWGIVPLSGKSLPDSNSES
IDNPEFALLFNLAARAATLSAPGPFNFGNGKGNKKNPNLNOSNKKPSNP
NOMKKGKGNSSHEAMNEMNGOCPLARSYAMSGILPLVAKALQPPAGVKLKC
PPAVVLAARALTELKSLRPPQPAKVAIALLGNAIPLGVWREHTKFSQWPF
AAVHAAPVFFGLMKLSVNMPTANAFITAAISIIQITIGSPRAERIRKALAAGDADST
TVADMYPNKSGNCSDETEGKAWDPLAMKMGASGVAAVTPSNCF"

gene 30681..31229
/gene="OJ1714_H10.11"
/db_xref="GI:22831074"
/translation="MSAGSPSPGGGSGGSGIFSGGIGGPGYGVGASVGVLLVUS
TVLAITFCSTNPFVAAAGRPFPFRPDGDDVIGIGIDATLEAFPEVYGBA
RKAGGAATCTCPVLENGYDGVLRALPDGHLFRCVDFWLQRTPCVCRTS
PLSPMPTPLAETPLALVRES"
complement(join(32159..32488,32607..32675,32753..32842,
32995..33060,33155..33205,33327..33584,33875..33916))
/gene="OJ1714_H10.12"
complement(join(32159..32488,32607..32675,32753..32842,
32995..33060,33155..33205,33327..33584,33875..33916))
/gene="OJ1714_H10.12"
/note="somatic embryogenesis receptor-like kinase-like protein"

protein
/codon_start=1
/protein_id="BAC15937.1"
/db_xref="GI:22831075"
/translation="MGCLSRHQLCGGNGLPFLKPKHKEATNGFSALLRAGPGG
RAAYRARPADGLVATVRAGDGDQDREAFYRELQLLARLNHRHIVLHYSQGHFL
VFDQMRSLKECHUPPLNWRTRICVAIDVAALLEYLYPCDPVPLVHVSNSNP
MFMENAKLSIDVIGDPRKTVESNAASPEIQORRLDVFQYVLIILEVTGOSP
GEGELVQVQEPGACTMYXVDADLNTYDSKELRNLIARLCTRPGNDAMVSI
LILYLOQKVANLGCSENICLSV"
join(34057..34198,34299..34362,34541..34616)
/gene="OJ1714_H10.13"
join(34057..34198,34299..34362,34541..34616)
/gene="OJ1714_H10.13"
/note="This category is not included in IRGSP standard.
hypothetical ORF

Query Match 6.7%; Score 45; DB 8; Length 191580;
Best Local Similarity 100.0%; Pred. No. 1e-13;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 135 AACATGCCAGACTGCCATGCATTCACCATAGCAGCTCCATT 179
Db 25704 AACATGCCAGACTGCCATGCATTCACCATAGCAGCTCCATT 25660

RESULT 6
AL773560/c
LOCUS 108859 bp DNA linear MAM 02-OCT-2002
DEFINITION Pig DNA sequence from clone XX-514B12, complete sequence.
ACCESSION AL773560
VERSION AL773560.8 GI:23497928
KEYWORDS HTG.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
Sinha, H.
Direct Submission
Submitted (02-OCT-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk

chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep. This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr13>. RP11-184L18 is from the library RPC1-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBAC3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-184L18. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RP11-184L18 is at 1 in this sequence. The true left end of clone RP11-103 is at 121397 in this sequence. The true right end of clone RP11-150L15 is at 45061 in this sequence.

FEATURES

Location/Qualifiers

```

source
1..121496
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-184L18"
/clone_lib="RPC1-11.1"
80..135
/note="28 copies 2 mer at 71% conserved"
repeat_region
809..1157
/note="MTLJ repeat: matches 37..408 of consensus"
repeat_region
2725..2774
/note="25 copies 2 mer tt 76% conserved"
repeat_region
2806..3198
/note="12 repeat: matches 2323..2691 of consensus"
repeat_region
3314..3377
/note="32 copies 2 mer tt 67% conserved"
repeat_region
3698..4004
/note="HAL1 repeat: matches 569..877 of consensus"
repeat_region
4093..4152
/note="15 copies 4 mer atgt 73% conserved"
repeat_region
4283..4513
/note="LIM2 repeat: matches 5379..5611 of consensus"
repeat_region
4533..4723
/note="LIM9 repeat: matches 5975..6177 of consensus"
repeat_region
4724..5039
/note="ALUS9 repeat: matches 1..312 of consensus"
repeat_region
5044..5085
/note="21 copies 2 mer aa 81% conserved"
repeat_region
5090..5376
/note="ALUSJ repeat: matches 1..287 of consensus"
repeat_region
5377..5412
/note="12 copies 3 mer taa 97% conserved"
repeat_region
5422..5573
/note="LIM9 repeat: matches 6155..6308 of consensus"
repeat_region
6247..6410
/note="MER53 repeat: matches 1..173 of consensus"
repeat_region
6609..6900
/note="ALUSG repeat: matches 1..293 of consensus"
repeat_region
9635..9690
/note="22 copies 3 mer gtg 83% conserved"
repeat_region
10471..10803
/note="LIM8 repeat: matches 5959..6291 of consensus"
repeat_region
12037..12100
/note="32 copies 2 mer tc 76% conserved"
repeat_region
12129..14123
/note="LINEC repeat: matches 1536..3241 of consensus"

```

```

repeat_region
14124..14489
/note="THE1B repeat: matches 1..364 of consensus"
repeat_region
14490..14618
/note="LINEC repeat: matches 1411..1536 of consensus"
repeat_region
14693..14857
/note="WTLIP repeat: matches 357..513 of consensus"
repeat_region
16039..16285
/note="LIM2 repeat: matches -569..-420 of consensus"
repeat_region
16286..16832
/note="MER41A repeat: matches 3..554 of consensus"
repeat_region
16833..16996
/note="LIM2 repeat: matches -421..-258 of consensus"
repeat_region
17440..17546
/note="MER41B repeat: matches 410..518 of consensus"
repeat_region
17895..17963
/note="NST-INTERNAL repeat: matches 748..827 of consensus"
repeat_region
18051..18349
/note="LIPBA repeat: matches 64..369 of consensus"
repeat_region
18524..19831
/note="LIPBA repeat: matches 383..1687 of consensus"
repeat_region
20184..21299
/note="MER52A repeat: matches 660..1755 of consensus"
repeat_region
21381..21667
/note="LIM37 repeat: matches 5973..6262 of consensus"
repeat_region
21892..22426
/note="LIMB1 repeat: matches 5617..6158 of consensus"
repeat_region
22423..22597
/note="LIMB6 repeat: matches 5664..5839 of consensus"
repeat_region
22616..22814
/note="LIM4 repeat: matches 3081..3289 of consensus"
repeat_region
22841..23223
/note="WTLAL repeat: matches 1..365 of consensus"
repeat_region
23337..23844
/note="LI repeat: matches 2264..2556 of consensus"
repeat_region
23647..24071
/note="LIM4C repeat: matches 264..690 of consensus"
repeat_region
24187..24207
/note="MER51-internal repeat: matches 7314..7333 of consensus"
repeat_region
24208..24296
/note="MER57-internal repeat: matches 7151..7244 of consensus"
repeat_region
24958..25001
/note="11 copies 4 mer aagg 79% conserved"
repeat_region
26298..26349
/note="12 repeat: matches 2685..2743 of consensus"
repeat_region
27194..27435
/note="LIM4 repeat: matches 5339..5607 of consensus"
repeat_region
27428..27778
/note="LIM4 repeat: matches 4329..4691 of consensus"
repeat_region
27789..27849
/note="LIM4 repeat: matches 5731..5786 of consensus"
repeat_region
27948..28034
/note="LIM4 repeat: matches 5183..5271 of consensus"
repeat_region
28070..28651
/note="LI repeat: matches 3508..4119 of consensus"
repeat_region
28761..29156
/note="LIM10 repeat: matches 5887..6296 of consensus"
repeat_region
29294..29341
/note="24 copies 2 mer tt 75% conserved"
repeat_region
29694..30004
/note="ALUY repeat: matches 1..309 of consensus"
repeat_region
30024..30107
/note="12 repeat: matches 2406..2489 of consensus"
repeat_region
30564..30849
/note="ALUSQ repeat: matches 1..286 of consensus"
repeat_region
31135..31423
/note="ALUSX repeat: matches 26..311 of consensus"
repeat_region
34487..34500
/note="LIM8 repeat: matches 6099..6284 of consensus"
repeat_region
34931..35866
/note="MERV16 repeat: matches 10..941 of consensus"
repeat_region
35924..37229

```

repeat_region /note="THE1B-INTERNAL repeat: matches 5. .1340 of consensus" 37236. 37584

repeat_region /note="THE1B repeat: matches 3. .362 of consensus" 38216. 38279

repeat_region /note="16 copies 4 mer tata 68% conserved" 38224. 38277

repeat_region /note="27 copies 2 mer ta 74% conserved" 38472. 38859

repeat_region /note="194 copies 2 mer tt 56% conserved" 40595. 41191

repeat_region /note="L1MB4 repeat: matches 5568. .6174 of consensus" 44397. 44652

repeat_region /note="AluB repeat: matches 39. .183 of consensus" 45462. 45565

repeat_region /note="L2 repeat: matches 2640. .2742 of consensus" 47671. 48031

repeat_region /note="THE1B repeat: matches 1. .364 of consensus" 48599. 48910

repeat_region /note="AluSP repeat: matches 1. .313 of consensus" 49131. 49259

repeat_region /note="L2 repeat: matches 2142. .2274 of consensus" 49405. 49714

repeat_region /note="AluSg repeat: matches 1. .308 of consensus" 50255. 50466

repeat_region /note="MER20 repeat: matches 2. .218 of consensus" 50886. 51111

repeat_region /note="MIR repeat: matches 5. .234 of consensus" 51180. 51597

Query Match 3.1%; Score 21; DB 9; Length 121496;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 651 TATCTGTGTGTAAAAAAA 671

DB 100692 TATCTGTGTGTAAAAAAA 100672

RESULT 8

AC108178

LOCUS AC108178 145257 bp DNA linear HTG 11-SEP-2002

DEFINITION Felis catus clone RP86-113N11, WORKING DRAFT SEQUENCE, 2 ordered pieces

ACCESSION AC108178

VERSION AC108178.2 GI:22779523

KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.

SOURCE Felis catus (cat)

ORGANISM Felis catus

REFERENCE 1 (bases 1 to 145257)

AUTHORS Akhter N., Antonellis A., Ayele K., Beckstrom-Sternberg S.M., Bukaryota, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis. Benjamin B., Blakesley R.W., Bouffard G.G., Breen K., Brinkley C., Brooks S., Dietrich N.L., Granite S., Guan X., Gupta J., Haghighi P., Han J., Hansen N., Ho S.-L., Idol J.R., Karling B., Laric P., Lee-Lin S.-Q., Legaapi R., Maduro Q.L., Maduro V.B., Margulies E.H., Mastello C., Maskeri B., Mastrian S.D., McCloskey J.C., McDowell J., Pagnuirigan C., Pearson R., Portnoy M.B., Prasad A., Reddix-Dugue N., Schueler M.G., Sison C., Stantrypop S., Thomas J.W., Thomas P.J., Touchman J.W., Vogt J.L., Wetherby K.D., Wiggins L., Young A., Zhang L.-H. and Green E.D.

TITLE NISC Comparative Sequencing Initiative

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 145257)

AUTHORS Green E.D.

TITLE Direct Submission

JOURNAL Submitted (26-JAN-2002) NIH Intramural Sequencing Center, 8717 Government Circle, Gaithersburg, MD 20877, USA

REFERENCE 3 (bases 1 to 145257)

AUTHORS Green E.D.

TITLE Direct Submission

JOURNAL Submitted (11-SEP-2002) NIH Intramural Sequencing Center, 8717

COMMENT

Grovemont Circle, Gaithersburg, MD 20877, USA
On Sep 11, 2002 this sequence version replaced gi:19376877.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.zoengri.nih.gov
----- Project Information
Center project name: cfq
Center clone name: 113N11

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8x average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 144088 bases at least Q40
Consensus quality: 144686 bases at least Q30
Consensus quality: 145012 bases at least Q20
Insert size: 117000; agarose-fp
Insert size: 145157; sum-of-contigs
Quality coverage: 10.38x in Q20 bases; agarose-fp
Quality coverage: 8.37x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 93196: contig of 93196 bp in length
* 93197 93296: gap of unknown length
* 93297 145257: contig of 51961 bp in length.
* Location/Qualifiers
1..145257
/organism="Felis catus"
/mol_type="genomic DNA"
/db_xref="taxon:9685"
/clone="RP86-113N11"
/clone_lib="RP86"
1..93196
/note="assembly_fragment
clone_end:SP6
vector_side:left"
1..145257
/note="clone overlaps with GenBank Accession Number
AC108188 clone RP86-138N10 (center project name cfz)"
93297..145257
/note="assembly_fragment
clone_end:T7
vector_side:right"

BASE COUNT 44526 a 25773 g 48227 t 100 others

ORIGIN

Query Match 3.1%; Score 21; DB 2; Length 145257;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 410 TTGGAAATTTCTATCTACCA 430

Db 113983 TTGGAAATTTGTATCTACCA 114003

|||||

RESULT 9
AC021868
LOCUS
DEFINITION
AC021868 Homo sapiens 12 BAC RP11-679N16 (Roswell Park Cancer Institute Human BAC Library) complete sequence.
AC021868
VERSION
KEYWORDS
SOURCE
ORGANISM
HTG. AC021868.17 GI:11496284 Homo sapiens (human)

REFERENCE
AUTHORS
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C., Albrooks, S.L., Amarantunga, H.C., Are, J.R., Banks, T., Barbaria, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chiu, D., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dethorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Emerling, S., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Han, J., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B., Honsi, F., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Kratochvic, J., Kureshi, A., Landry, N., Leal, B., Lee, E., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulesed, H., Lorado, R., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maneshwari, M., Mapua, P., Marondel, I., Martin, R., Martindale, A., Martinez, E., Massey, S., Mawhinney, S., McLeod, M.P., Meador, M., Mei, G., Merscher, S., Metzger, M., Miller, A., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Montgomery, K.T., Morgan, M., Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokwenkwo, S., Ogutu, M., Okwodu, G., Oraguine, N., Oviedo, R., Pace, A., Payton, B., Perry, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Qules, M., Ren, Y., Rivas, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shim, C., Shooshari, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Swatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Zucheraipati, R. and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 168227)
Worley, K.C.
Direct Submission
Submitted (21-JAN-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 168227)
Worley, K.C.
Direct Submission
Submitted (01-DEC-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Dec 1, 2000 this sequence version replaced gi:10716564.

INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished) for Human and Mouse sequences.
Genes and region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and CDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

Contig length: 168227
Phrap values in estimate: 167445
Average error rate (BCM-Phrap estimate): 0.000356395
Fraction of Phrap values less than 40: 0.0300815
Number of consensus changing edits: 14
Number of N's in consensus: 0

----- Summary Statistics -----

Position	Original-Context	Edited-Context
109003	gcccagctcg(n)atatttct	gcccagctcg(g)atatttct
134203	ccaaataata(n)agacaactaa	ccaaataata(a)agacaactaa
134323	tttgggtcca(n)atctgggtt	tttgggtcca(a)atctgggtt
149014	ttttttttt(a)agnatagacg	ttttttttt(t)agnatagacg
163499	gcagaaatta(n)caaatgacg	gcagaaatta(a)caaatgacg
164706	agctcccgga(n)taantgggat	agctcccgga(g)tagctgggat
164709	ctcccganta(n)ntgggattac	ctcccganta(g)ctgggattac
164710	tcccgantan(n)tggaattaca	tcccganta(g)ctgggattac
164869	gttgatttctg(n)ctgctnnnn	gttgatttctg(g)ctgctnnnn
164874	ttgctgctg(n)nnnnccaaa	ttgctgctg(c)ctgctnnnn
164875	ttgctgctg(n)nnnnccaaa	ttgctgctg(c)ctgctnnnn
164876	gtctgctg(n)nnnnccaaa	gtctgctg(c)ctgctnnnn
164878	gtctgctg(n)nnnnccaaa	gtctgctg(c)ctgctnnnn
166339	gggtaagggg(n)ttcttcttgg	gggtaagggg(t)ttcttcttgg

----- Distribution of Quality < 40 Bases -----

#	1000	900	800	700
*	*	*	*	*
*	*	*	*	*
*	*	*	*	*
*	*	*	*	*

```

bases 600 | * * * *
       500 | * * * *
       400 | * * * *
       300 | * * * *
       200 | * * * *
       100 | * * * *
         0 | * * * *
-----
          5 10 15 20 25 30 35 40
          Phrap Value Range

```

```

FEATURES
source
  Version: 1.01 qxf0.
  Location/Qualifiers
    1..168227
      /organism="Homo sapiens"
      /mol_type="genomic DNA"
      /db_xref="taxon:9606"
      /chromosome="12"
      /clone="RP11-679N16"
      375..520
        /rpt_family="MIR"
      521..798
        /rpt_family="AluJo"
      799..937
        /rpt_family="MIR"
      1474..1501
        /rpt_family="AT rich"
        complement(1759..1975)
      /rpt_family="LIMB8"
      complement(1976..2341)
      /rpt_family="AluJb"
      complement(2342..2440)
      /rpt_family="LIMB8"
      complement(2441..2784)
      /rpt_family="LIP4"
      complement(2785..2957)
      /rpt_family="LIMB8"
      complement(2962..3137)
      /rpt_family="AluJb"
      3138..3172
        /rpt_family="(TTTG)n"
        complement(3173..3320)
      /rpt_family="AluJb"
      3367..3661
        /rpt_family="AluSq"
        complement(3677..3726)
      /rpt_family="LIM4"
      complement(3768..4362)
      /rpt_family="LIM4"
      4366..4418
        /rpt_family="(TA)n"
        complement(4419..4702)
      /rpt_family="AluSx"
      complement(4703..4747)
      /rpt_family="LIM4"
      complement(4812..5353)
      /rpt_family="LIM4"
      complement(5447..6000)

```

```

Query Match      3.1%; Score 21; DB 9; Length 168227;
Best Local Similarity 100.0%; Pred.No. 4.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 410 TTGGAAATTTCTATCTACCA 430
Db 160166 TTGGAAATTTCTATCTACCA 160186

```

```

RESULT 10
AC019209/c AC019209 196817 bp DNA linear PRI 07-OCT-2000
LOCUS

```

```

DEFINITION
Homo sapiens BAC clone RP11-443N24 from Sequence-12, complete
sequence.
ACCESSION
AC019209
VERSION
AC019209.3 GI:7684571
KEYWORDS
HTG.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 196817)
Sulston,J.E. and Waterston,R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
9847074
2 (bases 1 to 196817)
Palk,A., Maupin,R. and Du,H.
The sequence of Homo sapiens BAC clone RP11-443N24
Unpublished
3 (bases 1 to 196817)
Waterston,R.H.
Direct Submission
Submitted (30-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 196817)
Waterston,R.H.
Direct Submission
Submitted (12-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 196817)
Waterston,R.H.
Direct Submission
Submitted (07-OCT-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 3, 2000 this sequence version replaced gi:6691391.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics
-----
Center project name: H_NH0443N24
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e. phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male

donor, as described by Osoegawa, K., Woon, P. Y., Zhao, B., Frengen, B., Tateno, M., Catanese, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

NEIGHBORING SEQUENCES INFORMATION:

Actual start of this clone is at base position 1 of RP11-443N24; actual end is at base position 196817 of RP11-443N24.

FEATURES

```

source
1..196817
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="12"
/map="Sequence-12"
/clone="RP11-443N24"
/clone_lib="RC11-11"
837..925
/rpt_family="MER1_type"
repeat_region
847..928
/rpt_family="MER1_type"
repeat_region
925..951
/rpt_family="MER1_type"
repeat_region
952..1467
/rpt_family="Retroviral"
repeat_region
1468..1518
/rpt_family="MER1_type"
repeat_region
1573..1626
/rpt_family="L2"
repeat_region
1803..1878
/rpt_family="MIR"
repeat_region
2050..2259
/rpt_family="MIR"
repeat_region
2265..2649
/rpt_family="MER2_type"
repeat_region
2928..3004
/rpt_family="MER2_type"
repeat_region
3005..3091
/rpt_family="MER2_type"
repeat_region
3145..3159
/rpt_family="L2"
repeat_region
3160..3183
/rpt_family="(TTA)n"
repeat_region
3184..3247
/rpt_family="L2"
repeat_region
4744..5058
/rpt_family="MaLR"
repeat_region
5100..5182
/rpt_family="MER113"
repeat_region
6049..6480
/rpt_family="Retroviral"
repeat_region
6497..6784
/rpt_family="Alu"
repeat_region
6786..6814
/rpt_family="AT-rich"
repeat_region
9126..9199
/rpt_family="L1"
repeat_region
9200..9495
/rpt_family="Alu"
repeat_region
9496..9757
/rpt_family="L1"
repeat_region
9817..10132
/rpt_family="Alu"
repeat_region
10133..11197
/rpt_family="L1"
repeat_region
11235..11318
/rpt_family="L1"
repeat_region
11367..11399
/rpt_family="AT-rich"
repeat_region
11770..12111

```

```

repeat_region
12200..12463
/rpt_family="MaLR"
repeat_region
12490..12634
/rpt_family="L2"
repeat_region
12788..12842
/rpt_family="T-rich"
repeat_region
12919..12984
/rpt_family="L2"
repeat_region
13385..13487
/rpt_family="MER89"
repeat_region
13517..13960
/rpt_family="MER89"
repeat_region
14060..14249
/rpt_family="MIR"
repeat_region
14317..14743
/rpt_family="MaLR"
repeat_region
14748..16261
/rpt_family="MaLR"
repeat_region
16310..16433
/rpt_family="MaLR"
repeat_region
16432..16606
/rpt_family="MaLR"
repeat_region
16831..16885
/rpt_family="(CA)n"
repeat_region
17239..17472
/rpt_family="MIR"
repeat_region
17832..17898
/rpt_family="AT-rich"
repeat_region
19263..19709
/rpt_family="Retroviral"
repeat_region
20096..20152
/rpt_family="L2"
repeat_region
20104..20207
/rpt_family="MIR"
repeat_region
20208..20514
/rpt_family="Alu"
repeat_region
20515..20636
/rpt_family="MIR"
repeat_region
21061..21230
/rpt_family="L2"
repeat_region
21331..21651
/rpt_family="Retroviral"
repeat_region
21661..22248
/rpt_family="MER1_type"
repeat_region
22272..22350
/rpt_family="MER1_type"
repeat_region
22349..22432
/rpt_family="MER1_type"
repeat_region
22434..22590
/rpt_family="MER1_type"
repeat_region
23549..23695
/rpt_family="L2"

```

Query Match 3.1%, Score 21; DB 9; Length 196817;

Best Local Similarity 100.0%; Pred.No. 4.2;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 410 TTGGAAATTTGTATCTACCA 430

Db 22146 TTGGAAATTTGTATCTACCA 22126

RESULT 11

AC044811

LOCUS

DEFINITION Homo sapiens clone RP11-522D14, WORKING DRAFT SEQUENCE, 25

AC044811 205992 bp DNA linear HTG 21-MAY-2000

VERSION AC044811.2 GI:7960389

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 205992)
Birren,B., Linton,L., Nubaum,C., Lander,B., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bida,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Collangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
Dodds,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galgani,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheters,R.,
Melarim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Ollivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tittell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (12-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 21, 2000 this sequence version replaced gi:7543770.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7086
Center clone name: 522.D.14
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 190410 bases at least Q40
Consensus quality: 198607 bases at least Q30
Consensus quality: 201747 bases at least Q20
Insert size: 198000; agarose-1p
Insert size: 203592; sum-of-contigs
Quality coverage: 4.4 in Q20 bases; agarose-1p
Quality coverage: 4.3 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1395: contig of 1395 bp in length
* 1396: gap of 100 bp
* 1496: contig of 1862 bp in length
* 3457: gap of 100 bp
* 5037: contig of 1580 bp in length
* 5038: gap of 100 bp
* 5138: contig of 2263 bp in length
* 7401: gap of 100 bp
* 7501: 9999: contig of 2399 bp in length
* 9900: gap of 100 bp
*
* 10000: 12544: contig of 2545 bp in length
* 12545: gap of 100 bp
* 12645: contig of 1380 bp in length
* 14024: gap of 100 bp
* 14025: contig of 3712 bp in length
* 17837: gap of 100 bp
* 17937: 21475: contig of 3539 bp in length
* 21476: gap of 100 bp
* 21576: contig of 3069 bp in length
* 24645: gap of 100 bp
* 24646: contig of 3069 bp in length
* 27444: gap of 100 bp
* 28004: contig of 5060 bp in length
* 29805: gap of 100 bp
* 34386: contig of 4482 bp in length
* 34387: gap of 100 bp
* 34486: contig of 4778 bp in length
* 39265: gap of 100 bp
* 39266: contig of 4778 bp in length
* 43398: gap of 100 bp
* 43399: contig of 6034 bp in length
* 45398: gap of 100 bp
* 45399: contig of 9372 bp in length
* 54870: contig of 9372 bp in length
* 54971: gap of 100 bp
* 61581: contig of 6611 bp in length
* 61582: gap of 100 bp
* 61692: contig of 8046 bp in length
* 69827: gap of 100 bp
* 69828: contig of 7259 bp in length
* 77086: gap of 100 bp
* 77087: contig of 9375 bp in length
* 86561: gap of 100 bp
* 86562: contig of 10783 bp in length
* 97444: gap of 100 bp
* 97445: contig of 14208 bp in length
* 111753: gap of 100 bp
* 111754: contig of 14208 bp in length
* 126521: gap of 100 bp
* 126522: contig of 15385 bp in length
* 142007: gap of 100 bp
* 142008: contig of 31134 bp in length
* 173241: gap of 100 bp
* 173242: contig of 32652 bp in length.
*
FEATURES
Location/Qualifiers
1..205992
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="RP11-522D14"
/clone_lib="RPC1-11 Human Male BAC"
1..1395
/note="assembly_fragment"
1496..3357
/note="assembly_fragment"
3458..5037
/note="assembly_fragment"
5138..7400
/note="assembly_fragment"
7501..9899
/note="assembly_fragment"
10000..12544
/note="assembly_fragment"
12645..14024
/note="assembly_fragment"
clone_end:SP6
vector_side:left
14125..17836
/note="assembly_fragment"
17937..21475
/note="assembly_fragment"
21576..24644
/note="assembly_fragment"
24745..29804
/note="assembly_fragment"
29905..34386
/note="assembly_fragment"
34487..39264
/note="assembly_fragment"

```

/note="assembly_fragment"
39365..45398
/note="assembly_fragment"
45499..54870
/note="assembly_fragment"
54971..61581
/note="assembly_fragment"
61682..69727
/note="assembly_fragment"
69838..77086
/note="assembly_fragment"
77187..86561
/note="assembly_fragment"
86662..97444
/note="assembly_fragment"
97545..111753
/note="assembly_fragment"
111853..126521
/note="assembly_fragment"
126622..142006
/note="assembly_fragment"
142107..173240
/note="assembly_fragment"
clone_end.T7
vector_side:left
173341..205992
/note="assembly_fragment"
39890 c 38789 g 61370 t 2402 others
BASE COUNT 63541 a 39890 c 38789 g 61370 t 2402 others
ORIGIN

Query Match
Best Local Similarity 100.08; Pred.No. 4.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 410 TTGGAATTTGTATCTACCA 430
DB 17480 TTGGAATTTGTATCTACCA 17500

RESULT 12
AC109203/c
LOCUS 211299 bp DNA linear HTG 22-MAY-2003
DEFINITION Mus musculus clone RP23-2B18, *** SEQUENCING IN PROGRESS ***, 11
unordered pieces.
ACCESSION AC109203
VERSION AC109203.5 GI:30985030
KEYWORDS HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 211299)
Birren, B., Nusbaum, C. and Lander, E.
Mus musculus, clone RP23-2B18
Unpublished
2 (bases 1 to 211299)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouckghalter, B.,
Brown, A., Canarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Choepe, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fero, S.,
Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kanat, A., Karatas, A., Kella, C., LaRocque, K., Lamazares, R.,
Lander, E., Lechoczky, J., Levine, R., Liu, G., MacLean, C.,
Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,
McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T.,
Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S.,
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (03-FEB-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 211299)
Birren, B., Nusbaum, C., Lander, E., Abouelheil, A., Allen, N.,
Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,
Boguslavsky, L., Bouckghalter, B., Camarata, J., Chang, J., Choepe, Y.,
Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K.,
Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S.,
Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S.,
Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kanat, A., Karatas, A., Kella, C., Lander, E., Levine, R.,
Lindblad-Toh, K., Liu, G., Liu, A., Mabbitt, R., Maclean, C.,
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J.,
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C.,
Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,
Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,
Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (22-MAY-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 22, 2003 this sequence version replaced gi:30270695.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIGR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submission@genome.wi.mit.edu
----- Project Information
Center project name: L20517
Center clone name: 2_B_18
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 31141: contig of 31141 bp in length
* 31142 31241: gap of 100 bp
* 31242 74433: contig of 43192 bp in length
* 74434 74533: gap of 100 bp
* 74534 77038: contig of 2505 bp in length
* 77039 77138: gap of 100 bp
* 77139 81439: contig of 4301 bp in length
* 81440 81539: gap of 100 bp
* 81540 94256: contig of 12717 bp in length
* 94257 94356: gap of 100 bp
* 94357 100133: contig of 5777 bp in length
* 100134 100233: gap of 100 bp
* 100234 107322: contig of 7094 bp in length
* 107323 107427: gap of 100 bp
* 107428 149160: contig of 41733 bp in length
* 149161 149260: gap of 100 bp
* 149261 165074: contig of 15814 bp in length
* 165075 165174: gap of 100 bp
* 165175 193176: contig of 28002 bp in length
* 193177 193276: gap of 100 bp

```

```

Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (03-FEB-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 211299)
Birren, B., Nusbaum, C., Lander, E., Abouelheil, A., Allen, N.,
Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,
Boguslavsky, L., Bouckghalter, B., Camarata, J., Chang, J., Choepe, Y.,
Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K.,
Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S.,
Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S.,
Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kanat, A., Karatas, A., Kella, C., Lander, E., Levine, R.,
Lindblad-Toh, K., Liu, G., Liu, A., Mabbitt, R., Maclean, C.,
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J.,
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C.,
Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,
Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,
Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (22-MAY-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 22, 2003 this sequence version replaced gi:30270695.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIGR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submission@genome.wi.mit.edu
----- Project Information
Center project name: L20517
Center clone name: 2_B_18
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 31141: contig of 31141 bp in length
* 31142 31241: gap of 100 bp
* 31242 74433: contig of 43192 bp in length
* 74434 74533: gap of 100 bp
* 74534 77038: contig of 2505 bp in length
* 77039 77138: gap of 100 bp
* 77139 81439: contig of 4301 bp in length
* 81440 81539: gap of 100 bp
* 81540 94256: contig of 12717 bp in length
* 94257 94356: gap of 100 bp
* 94357 100133: contig of 5777 bp in length
* 100134 100233: gap of 100 bp
* 100234 107322: contig of 7094 bp in length
* 107323 107427: gap of 100 bp
* 107428 149160: contig of 41733 bp in length
* 149161 149260: gap of 100 bp
* 149261 165074: contig of 15814 bp in length
* 165075 165174: gap of 100 bp
* 165175 193176: contig of 28002 bp in length
* 193177 193276: gap of 100 bp

```



```

/mol_type="genomic DNA"
/db_xref="taxon:10116"
/c1one="CH230-254F13"
misc_feature
1375..2296
/note="clone boundary"
clone_end=T7
site=Nbcl
end sequence:RXACG31TJ"
BASE COUNT 56129 a 47161 c 48018 g 56496 t 10003 others
ORIGIN

Query Match 3.1%; Score 21; DB 2; Length 217807;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 465 TATTATTCGTGATGATTAAT 485
DB 29862 TATTATTCGTGATGATTAAT 29882

RESULT 14
AC122944
LOCUS Rattus norvegicus clone CH230-4F17, *** SEQUENCING IN PROGRESS ***
DEFINITION AC122944
VERSION AC122944.4 GI:30466953
KEYWORDS HTG; HTGS PHAS31; HTGS DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 249358)
Muzny,D.,Marle., Metzker,M.,Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsebrook,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Ayoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,P.,
Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,P., Garcia,A., Garner,T., Garza,M.,
Hollins,B., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kwis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,M., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuewa,L., Louised,H., Lozada,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartine,M., Mahmood,M., Malloy,K., Mangum,A.,
Mangum,B., Mapus,P., Martin,K., Martin,R., Martinez,E.,
Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Narkervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plopper,P., Poldexter,A., Popovic,D., Primus,E., Fu,L.,
Puzo,M., Quiroz,J., Rachlin,B., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savory,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umami,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleciyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 248358)
Worley,K.C.
Direct Submission
Submitted (26-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 248358)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 9, 2003 this sequence version replaced gi:23265699.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GAPX
Center clone name: CH230-4F17
----- Summary Statistics
Assembly program: Atlas
Consensus quality: 225281 bases at least Q40
Consensus quality: 229208 bases at least Q30
Consensus quality: 231758 bases at least Q20
Estimated insert size: 236500; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 246878: contig of 246878 bp in length
* 246879 246978: gap of unknown length
* 246979 248358: contig of 1380 bp in length.
Location/Qualifiers
1..248358
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-4F17"
1..2474
misc_feature

```

```

/note="wgs_end_extension
clone_end:T7"
/note="clone_boundary
clone_end:T7
site:EcoRI
end sequence: BH305410"
complement(243579..244808)
/note="clone_boundary
clone_end:Sp6
site:EcoRI
end sequence: BH305412"
BASE COUNT 62715 a 53910 c 54498 g 62596 t 14639 others
ORIGIN

Query Match 3.1% Score 21; DB 2; Length 248358;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 465 TATTTATCTTGATGTAAT 485
|||||
DB 216039 TATTTATCTTGATGTAAT 216059

RESULT 15
AC098275 254489 bp DNA linear HTG 10-OCT-2002
LOCUS Rattus norvegicus clone CH230-112, WORKING DRAFT SEQUENCE.
DEFINITION AC098275
ACCESSION AC098275
VERSION AC098275.4 GI:23664871
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 254489)
Muzny D.M., Adams C., Adio-Oduola B., Ali-Osman F.R., Allen C.,
Albrooks S.L., Amarantunga H.C., Are J.R., Ayele M., Banks T.,
Barbacia J., Benton J., Blum K., Blankenburg K., Bonnin D.,
Bouck J., Bowie S., Brileva M., Brown E., Brown M., Bryant N.P.,
Buhay C., Burch P., Burkett C., Burrell K.L., Byrd N.C.,
Carroll T.F., Carter M., Cavazos S.R., Chacko J., Chavez D.,
Chen G., Chen Z., Chen Z., Chowdhry I., Christopoulos C.,
Cleveland C.D., Cox C., Coyle M.D., Dathorne S.R., David R.,
Davila M.L., Davis C., Davy-Carroll L., Dederich D.A.,
Delaney K.R., Delgado O., Denn A.L., Ding Y., Dinh H.H.,
Douthwaite K.J., Draper H., Dugan-Rocha S., Durbin K.J.,
Earhart C., Edgar D., Edwards C.C., Elhaj C., Escotto M.,
Falls T., Ferraguto D., Flagg N., Ford J., Foster P., Frantz P.,
Gabisi A., Gao J., Garcia A., Garner T., Garza N., Gill R.,
Gorell J.H., Guevara W., Gunaratne P., Hale S., Hamilton K.,
Harris C., Harris K., Hart M., Havlak P., Hawes A., Hernandez J.,
Hernandez O., Hodgson A., Hogues M., Holloway C., Hollins B.,
Honsi F., Howard S., Huber J., Hulyk S., Hume J., Jackson L.E.,
Jacobson B., Jia Y., Johnson R., Jolivet S., Joudah S.,
Karissom B., Kelly S., Khan U., King L., Korvah J., Kovar C.,
Kratovic J., Kuresh A., Landry N., Leal B., Lewis L.C., Lewis L.,
Li J., Li Z., Lichtarge O., Lieu C., Liu J., Liu W., Loulseghe H.,
Lozato R.J., Lu X., Lucier A., Lucier R., Luna R., Ma J.,
Maheshwari M., Mapua P., Martin R., Martindale A., Martinez E.,
Massey E., Mathney E., McLeod M.P., Meador M., Mei G., Metzker M.,
Miner G., Miner Z., Mitchell T., Mohabbat N., Morgan M., Morris S.,
Mosier M., Neal D., Newton J., Newton B., Nguyen A., Nguyen N.,
Nguyen N., Nickerson E., Nwokwuo S., Oguh M., Okwuonu G.,
Oragunye N., Oviedo R., Pace A., Payton B., Peery J., Perez L.,
Peters L., Pickens R., Primus E., Pu L.L., Quiles M., Ren Y.,
Rives M., Rojas A., Rojibokan I., Rolfe M., Ruiz S., Savery G.,
Scherer S., Scott G., Shen H., Shoohtari N., Sisson I.,
Sodergren E., Sonaike T., Sparks A., Stanley H., Stone H.,
Sutton A., Svatek A., Tabor P., Tamerisa A., Tamerisa K., Tang H.,
Tansey J., Taylor C., Taylor T., Telford B., Thomas N., Thomas S.,
Usmani K., Vasquez L., Vera V., Villalon D., Vinson R., Wang Q.,

```

Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wlezyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.P., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G. and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 254489)
Worley, K.C.
Direct Submission
Submitted (23-OCT-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 254489)
Worley, K.C.
Direct Submission
Submitted (10-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Oct 10, 2002 this sequence version replaced gi:22002311. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc.help@bcm.tmc.edu
----- Project Information
Center project name: TURE
Center clone name: CH230-112
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 226734 bases at least Q40
Consensus quality: 230078 bases at least Q30
Consensus quality: 232204 bases at least Q20
Estimated insert size: 233069; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved
* 1 254489; contig of 254489 bp in length.
----- Location/Qualifiers
1. 254489
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-112"
1. 1712
/note="wgs_end_extension
clone_end:T7"
4870..6968
/note="wgs_end_extension
clone_end:T7"
complement(7119..7589)
/note="clone_boundary

FEATURES
source
misc_feature
misc_feature
misc_feature


```

clone_end:T7
site:EcoRI
end_sequence:BH296996"
complement(7127..7664)
/notes="clone_boundary
clone_end:T7
site:EcoRI
end_sequence:BH296996"
155350..156416
/notes="wgs_end_extension
clone_end:T7"
complement(250562..251206)
/notes="clone_boundary
clone_end:Sp6
site:EcoRI
end_sequence:BH296997"
251464..252618
/notes="wgs_end_extension
clone_end:T7"
253241..254489
/notes="wgs_end_extension
clone_end:T7"
BASE COUNT 62684 a 53733 c 53978 g 63097 t 20997 others
ORIGIN
Query Match 3.1%; Score 21; DB 2; Length 254489;
Best Local Similarity 100.0%; Fred. No. 4.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 465 TATTTATCTTGATGTAAT 485
Db 203263 TATTTATCTTGATGTAAT 203283

```

Search completed: November 14, 2003, 11:01:24
Job time : 1940 secs